

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 27, 2005, 20:17:15 ; Search time 268 Seconds
(without alignments)
8901.834 Million cell updates/sec

Title: US-10-829-107-3

Perfect score: 1458

Sequence: 1 gttgagagatgggatgtcc.....taataaaatgcactcattt 1458

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:*

- 1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
- 4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
- 5: /cgn2_6/ptodata/1/ina/PGTUS_COMB.seq.*
- 6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1458	100.0	1458	3	US-09-111-470-3
2	1458	100.0	1458	4	US-09-862-802A-3
3	1108	76.0	1370	3	US-09-111-470-9
4	1108	76.0	1370	4	US-09-862-802A-9
5	382	26.2	1224	4	US-09-949-016-4091
6	382	26.2	1277	4	US-09-016-434-1186
7	310.4	21.3	361	4	US-09-513-999C-13778
8	280.6	19.2	1348	4	US-09-949-016-4090
9	183	12.6	400	3	US-09-641-638-46
10	183	12.6	400	4	US-10-170-097-46
11	155.6	10.7	448	3	US-09-641-638-49
12	155.6	10.7	448	4	US-10-170-097-49
13	154.4	10.6	448	3	US-09-641-638-51
14	154.4	10.6	448	3	US-09-641-638-571
15	154.4	10.6	448	4	US-10-170-097-51
16	154.4	10.6	448	4	US-10-170-097-571
17	154	10.6	448	3	US-09-641-638-50
18	154	10.6	448	4	US-10-170-097-50
19	130.2	8.9	601	4	US-09-949-016-144698
20	130.2	8.9	601	4	US-09-949-016-144699
21	130.2	8.9	17723	4	US-09-949-016-15832
22	114.2	7.8	9740	4	US-09-949-016-15833
23	109.4	7.5	601	4	US-09-949-016-144713
24	76.4	5.2	601	4	US-09-949-016-144710
25	67.2	4.6	3259	5	PCT-US95-03747-1
26	61.4	4.2	2318	4	US-09-620-312D-733
27	57.4	3.9	997	4	US-09-907-794A-376

Sequence 376, App
Sequence 376, App
Sequence 376, App
Sequence 376, App
Sequence 376, App
Sequence 376, App
Sequence 376, App
Sequence 376, App
Sequence 10, Appl
Sequence 12, Appl
Sequence 13, Appl
Sequence 15, Appl
Sequence 16, Appl
Sequence 18, Appl
Sequence 19, Appl
Sequence 21, Appl
Sequence 4, Appl
Sequence 6, Appl

28 57.4 3.9 997 4 US-09-905-125A-376
29 57.4 3.9 997 4 US-09-902-775A-376
30 57.4 3.9 997 4 US-09-906-700-376
31 57.4 3.9 997 4 US-09-903-603A-376
32 57.4 3.9 997 4 US-09-904-920A-376
33 57.4 3.9 997 4 US-09-909-064-376
34 57.4 3.9 997 4 US-09-905-381A-376
35 57.4 3.9 997 4 US-09-906-618-376
36 53.8 3.7 417 3 US-09-535-521-10
37 53.8 3.7 417 3 US-09-535-521-12
38 53.8 3.7 423 3 US-09-535-521-13
39 53.8 3.7 423 3 US-09-535-521-15
40 53.8 3.7 561 3 US-09-535-521-16
41 53.8 3.7 561 3 US-09-535-521-18
42 53.8 3.7 624 3 US-09-535-521-19
43 53.8 3.7 624 3 US-09-535-521-21
44 53.8 3.7 876 3 US-09-535-521-4
45 53.8 3.7 876 3 US-09-535-521-6

ALIGNMENTS

RESULT 1
US-09-111-470-3
; Sequence 3, Application US/09111470
; Patent No. 6277959
; GENERAL INFORMATION:
; APPLICANT: Valladeau, Jenny
; APPLICANT: Ravel, Odile
; APPLICANT: Bates, Elizabeth E.M.
; APPLICANT: Ford, John
; APPLICANT: Saeland, Sem
; APPLICANT: Lebecque, Serge J.E.
; TITLE OF INVENTION: Mammalian Membrane Protein Genes;
; TITLE OF INVENTION: Related Reagents
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DNAX Research Institute
; STREET: 901 California Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1104

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/111,470
FILING DATE: 08-JUL-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/053,080
FILING DATE: 09-JUL-1997
ATTORNEY/AGENT INFORMATION:
NAME: Ching, Edwin P.
REGISTRATION NUMBER: 34,090
REFERENCE/DOCKET NUMBER: SF0695
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650)852-9196
TELEFAX: (650)496-1200
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1458 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 257..1204

```

;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 608
; OTHER INFORMATION: /note= "short form lacks
; OTHER INFORMATION: nucleotides 608-673"
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 775
; OTHER INFORMATION: /note= "ASGPRm (table 2) has
; OTHER INFORMATION: sequence insert encoding GRE between nucleotides 775-776"
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1064
; OTHER INFORMATION: /note= "nucleotide 1064 of DCMF2b
; OTHER INFORMATION: may be A, which would encode asn rather than asp at the resid
; OTHER INFORMATION: numbered 270"
;
US-09-111-470-3

Query Match      100.0%; Score 1458; DB 3; Length 1458;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1458; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTTGAGGAGATGGGATGTCACAGATGATAGGGCTCTGGGATTTTCAGACCCAAAGACGAC 60
Db 1 GTTGAGGAGATGGGATGTCACAGATGATAGGGCTCTGGGATTTTCAGACCCAAAGACGAC 60

Qy 61 AGGACTCCAGTCACTCTACCCAGCTCTCCAGGACACAGCGCTCCCAACTCTGAGTGAC 120
Db 61 AGGACTCCAGTCACTCTACCCAGCTCTCCAGGACACAGCGCTCCCAACTCTGAGTGAC 120

Qy 121 GTCCCACTCTGTGCTTTGAGCAGACAAACAAAGTGGGAATCACACCTCCAGACCTCCCA 180
Db 121 GTCCCACTCTGTGCTTTGAGCAGACAAACAAAGTGGGAATCACACCTCCAGACCTCCCA 180

Qy 181 CAGCTCACCCAGACATGGGCGCGGCTGCTCCATTTTCAGCTGTGACAACTCAGAG 240
Db 181 CAGCTCACCCAGACATGGGCGCGGCTGCTCCATTTTCAGCTGTGACAACTCAGAG 240

Qy 241 CCCTGTTGGCCCAAGCATGACAGGAGCTATGAAACTTCCAGTACTTGGAGAAATAGGT 300
Db 241 CCCTGTTGGCCCAAGCATGACAGGAGCTATGAAACTTCCAGTACTTGGAGAAATAGGT 300

Qy 301 GAAAGTCCAGGGTTTAAAAATGGGCCACTTCTCTCCAGTCCCTCCTGCAGCGTCTCCG 360
Db 301 GAAAGTCCAGGGTTTAAAAATGGGCCACTTCTCTCCAGTCCCTCCTGCAGCGTCTCCG 360

Qy 361 CTCTGGGCGCTGCATCTCTGCTGTCTGGGCGCTGGGCTGCTGCTGTGCTATCAT 420
Db 361 CTCTGGGCGCTGCATCTCTGCTGTCTGGGCGCTGGGCTGCTGCTGTGCTATCAT 420

Qy 421 CTGTGTGTTGGATTCCAAATTTCCAAATTTTCAGAGGGACCTGTGACCTCGAGAACAGA 480
Db 421 CTGTGTGTTGGATTCCAAATTTCCAAATTTTCAGAGGGACCTGTGACCTCGAGAACAGA 480

Qy 481 TTTTAGCACTTCACTCAACACTGTGGCGGAGATCCAGGCACTGACTTCCCAGGGCAG 540
Db 481 TTTTAGCACTTCACTCAACACTGTGGCGGAGATCCAGGCACTGACTTCCCAGGGCAG 540

Qy 541 CAGCTTGGAAAGAAACGATGATCTCTGAAAGCTGAGGTGGAGGGTTTCAAGCAGGAACG 600
Db 541 CAGCTTGGAAAGAAACGATGATCTCTGAAAGCTGAGGTGGAGGGTTTCAAGCAGGAACG 600

Qy 601 GCAGGCGGGGTATCTGAGCTCCAGGAACACATACGACAAAGGCACCTTAGGCCACTG 660
Db 601 GCAGGCGGGGTATCTGAGCTCCAGGAACACATACGACAAAGGCACCTTAGGCCACTG 660

Qy 661 TCCCACTGCCCCATCTGTGTGTGTCAGTTCATTTCTGAAATGCTCTCGAGTCCAGCA 720
Db 661 TCCCACTGCCCCATCTGTGTGTGTCAGTTCATTTCTGAAATGCTCTCGAGTCCAGCA 720

Qy 721 GCTGGTGCAGACCTGAGAAACTGACCTGCGAGGTGGCTACTCTCAACAAATGCTC 780
Db 721 GCTGGTGCAGACCTGAGAAACTGACCTGCGAGGTGGCTACTCTCAACAAATGCTC 780
```

```

Qy 781 CACTGAAGGGAAGCTGCTGCTCCCGCTCAACTGGGTGGAGACCAAGACAGAGCTGCTACTGGTT 840
Db 781 CACTGAAGGGAAGCTGCTGCTCCCGCTCAACTGGGTGGAGACCAAGACAGAGCTGCTACTGGTT 840

Qy 841 CTCTCACTCTGGGATGTCTGGGCCGAGGCTGAGAAAGTACTGCGCACTGAAAGAACGCCCA 900
Db 841 CTCTCACTCTGGGATGTCTGGGCCGAGGCTGAGAAAGTACTGCGCACTGAAAGAACGCCCA 900

Qy 901 CTTGGTGTGTCATCACTCCAGGGAGGAGCAGAAATTTTGTCCAGAAATATCTAGGCTCCGC 960
Db 901 CTTGGTGTGTCATCACTCCAGGGAGGAGCAGAAATTTTGTCCAGAAATATCTAGGCTCCGC 960

Qy 961 ATACACTCTGATGGGCTCAGTGACCTCAAGGAGGCTCGAAAGTGGTGGATGGAACAGA 1020
Db 961 ATACACTCTGATGGGCTCAGTGACCTCAAGGAGGCTCGAAAGTGGTGGATGGAACAGA 1020

Qy 1021 CTATGCGACCGGCTTCCAGAACTGGAAGCCAGGCCAGCAGACGACTGGCAGGGGACGG 1080
Db 1021 CTATGCGACCGGCTTCCAGAACTGGAAGCCAGGCCAGCAGACGACTGGCAGGGGACGG 1080

Qy 1081 GCTGGGTGAGGCGGAGGACTGTGCTCACTTCCATCCAGACGGCAGGTGGAATGACACGT 1140
Db 1081 GCTGGGTGAGGCGGAGGACTGTGCTCACTTCCATCCAGACGGCAGGTGGAATGACACGT 1140

Qy 1141 CTGCCAGAGGCCCTACCACTGGGTCTGCGAGGCTGGGCTGGGTGAGACCCAGGAGAG 1200
Db 1141 CTGCCAGAGGCCCTACCACTGGGTCTGCGAGGCTGGGCTGGGTGAGACCCAGGAGAG 1200

Qy 1201 TCACTGAGCTGCTTTGGTGGGACCAACCCGGGCCACAGAAATGGCGGTGGGAGGAGACTC 1260
Db 1201 TCACTGAGCTGCTTTGGTGGGACCAACCCGGGCCACAGAAATGGCGGTGGGAGGAGACTC 1260

Qy 1261 TTCTCAGGACCTCTCGCAAGACCGCTCTGGGAGAGAAATAGCACTGGGAGATTGGAAG 1320
Db 1261 TTCTCAGGACCTCTCGCAAGACCGCTCTGGGAGAGAAATAGCACTGGGAGATTGGAAG 1320

Qy 1321 CACTGTAAATTTTGAATTTTCTTTTAAATTTTAAAGATGTTAGTGTCTTA 1380
Db 1321 CACTGTAAATTTTGAATTTTCTTTTAAATTTTAAAGATGTTAGTGTCTTA 1380

Qy 1381 AGCTTTATTTTTTTTCCAACTTTTGAAGAGTCAACTTCATGAGGTATTAATTTTTTACATA 1440
Db 1381 AGCTTTATTTTTTTTCCAACTTTTGAAGAGTCAACTTCATGAGGTATTAATTTTTTACATA 1440

Qy 1441 ATAAAATGCACCTCATTT 1458
Db 1441 ATAAAATGCACCTCATTT 1458
```

RESULT 2

```

US-09-862-802A-3
; Sequence 3, Application US/09862802A
; Patent No. 6756478
; GENERAL INFORMATION:
; APPLICANT: Valladeau, Jenny
; APPLICANT: Ravel, Odile
; APPLICANT: Bates, Elizabeth Ester Mary
; APPLICANT: Ford, John
; APPLICANT: Lebecque, Serge J.E.
; APPLICANT: Saeland, Sem
; TITLE OF INVENTION: Isolated Mammalian Membrane Protein Genes; Related Reagents
; FILE REFERENCE: SF0695 B
; CURRENT APPLICATION NUMBER: US/09/862, 802A
; CURRENT FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/053,080
; PRIOR FILING DATE: 1997-07-09
; PRIOR APPLICATION NUMBER: US 09/111,470
; PRIOR FILING DATE: 1998-07-08
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 1458
```

; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: mammalian nucleic acid and protein
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (257)..(1204)
; OTHER INFORMATION: protein coding sequence
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (608)..(673)
; OTHER INFORMATION: short form lacks nucleotides 608-673
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (775)..(776)
; OTHER INFORMATION: ASGPRM (Table 2) has sequence insert encoding GBE between nucleot
; OTHER INFORMATION: than Asp at the residue numbered 270
; OTHER INFORMATION: 1064 of DMP2a may be A, which would encode Asn rather
; -09-862-802A-3

Query Match 100.0%; Score 1458; DB 4; Length 1458;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1458; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTTGAGGAGATGGGATGTCACAGATGATAGGGCTCTGGGATTTTCAGACCCCAAGACGAGC 60
DB 1 GTTGAGGAGATGGGATGTCACAGATGATAGGGCTCTGGGATTTTCAGACCCCAAGACGAGC 60
QY 61 AGGATCCAGTCACTTACCCAGCTCTCCAGGACACAGCGTCTCCAACTCTGAGTGAC 120
DB 61 AGGATCCAGTCACTTACCCAGCTCTCCAGGACACAGCGTCTCCAACTCTGAGTGAC 120
QY 121 GTCCACCTCTGCTCTGAGCAGACACCAACCTGGGAATCAGACCTCCAGACCTCCCA 180
DB 121 GTCCACCTCTGCTCTGAGCAGACACCAACCTGGGAATCAGACCTCCAGACCTCCCA 180
QY 181 CAGCTCCACCCAGACCTGGGCGCGGCTCCCTCCATTTTCAGCTGTGACAACTCAGAG 240
DB 181 CAGCTCCACCCAGACCTGGGCGCGGCTCCCTCCATTTTCAGCTGTGACAACTCAGAG 240
QY 241 CCCTGTTGGCCCAAGCATGCAAGGAGTATGAAACTTCCAGTACTTGGAGAATAAGGT 300
DB 241 CCCTGTTGGCCCAAGCATGCAAGGAGTATGAAACTTCCAGTACTTGGAGAATAAGGT 300
QY 301 GAAAGTCCAGGGGTTTAAATGAGGCACTTCTCTCCAGTCCCTCTGAGGCTCTCCG 360
DB 301 GAAAGTCCAGGGGTTTAAATGAGGCACTTCTCTCCAGTCCCTCTGAGGCTCTCCG 360
QY 361 CTCTGGGCTCTGCATCTCTGCTGTCTGGGCTCGGCTCTGCTGTGCTGTGCTCATCAT 420
DB 361 CTCTGGGCTCTGCATCTCTGCTGTCTGGGCTCGGCTCTGCTGTGCTGTGCTCATCAT 420
QY 421 CTCTGGGCTCTGCATCTCTGCTGTCTGGGCTCGGCTCTGCTGTGCTGTGCTCATCAT 480
DB 421 CTCTGGGCTCTGCATCTCTGCTGTCTGGGCTCGGCTCTGCTGTGCTGTGCTCATCAT 480
QY 481 TTTTAGCACTTCACTCAACACTGTGGCGAGATCCAGGCATGACTTCCAGGCGAG 540
DB 481 TTTTAGCACTTCACTCAACACTGTGGCGAGATCCAGGCATGACTTCCAGGCGAG 540
QY 541 CAGCTCGAAGAAACGATAGCATCTCTGAAAGCTGAGGTGGAGGTTTCAGCAGGAAACG 600
DB 541 CAGCTCGAAGAAACGATAGCATCTCTGAAAGCTGAGGTGGAGGTTTCAGCAGGAAACG 600
QY 601 GCAGGCGGGGTATCTGAGCTCCAGGAACACTTACCGAGAAGGCAACCTTAGGCCACTG 660
DB 601 GCAGGCGGGGTATCTGAGCTCCAGGAACACTTACCGAGAAGGCAACCTTAGGCCACTG 660
QY 661 TCCCACTGCCCATCTCTGTGTGTGCCAGGTTTCATTTCTGAAATGCTCTCTGCGAGTCCAGCA 720

DB 661 TCCCACTGCCCATCTCTGTGTGTGCCAGGTTTCATTTCTGAAATGCTCTCTGCGAGTCCAGCA 720
QY 721 GTTGTGCAAGACCTGGAAGAACTGACCTGCGGAGGCTGCTCTCAACAACTAGCTCTC 780
DB 721 GTTGTGCAAGACCTGGAAGAACTGACCTGCGGAGGCTGCTCTCAACAACTAGCTCTC 780
QY 781 CACTGAAGGACCTGCTGCGGCTCAACTGGGCTGAGGACCAAGACAGCTGCTACTGTT 840
DB 781 CACTGAAGGACCTGCTGCGGCTCAACTGGGCTGAGGACCAAGACAGCTGCTACTGTT 840
QY 841 CTCTCACTCTGGGATGCTCTGGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 900
DB 841 CTCTCACTCTGGGATGCTCTGGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 900
QY 901 CTCTGCTGCTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 960
DB 901 CTCTGCTGCTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 960
QY 961 ATACACTGGATGGGCTCACTGACCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 1020
DB 961 ATACACTGGATGGGCTCACTGACCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 1020
QY 1021 CTATGCGACCGGCTTCCAGAACTGGAAGCCAGGCCAGCAGCAGCTGGCAGGCGACCG 1080
DB 1021 CTATGCGACCGGCTTCCAGAACTGGAAGCCAGGCCAGCAGCAGCTGGCAGGCGACCG 1080
QY 1081 GCTGGGTGAGGCGGAGGCTGCTCACTTCCATCCAGACCGGCTGGAATGAGGCTGAGGCT 1140
DB 1081 GCTGGGTGAGGCGGAGGCTGCTCACTTCCATCCAGACCGGCTGGAATGAGGCTGAGGCT 1140
QY 1141 CTGCGAGGCGGCTTCCAGAACTGGAAGCCAGGCCAGCAGCAGCTGGCAGGCGACCG 1200
DB 1141 CTGCGAGGCGGCTTCCAGAACTGGAAGCCAGGCCAGCAGCAGCTGGCAGGCGACCG 1200
QY 1201 TCACTGAGCTGCTTGGTGGGACCACTGGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 1260
DB 1201 TCACTGAGCTGCTTGGTGGGACCACTGGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 1260
QY 1261 TTCTCAGACCTCTCTCGAAGACCGCTCTGGGAGAGAAATAAGCACTTGGAGATTTGGAAG 1320
DB 1261 TTCTCAGACCTCTCTCGAAGACCGCTCTGGGAGAGAAATAAGCACTTGGAGATTTGGAAG 1320
QY 1321 CACTGCTAATTTTGAATTTTCTCTTAAATTTTAAAGATGCTATAGTCTTCTTA 1380
DB 1321 CACTGCTAATTTTGAATTTTCTCTTAAATTTTAAAGATGCTATAGTCTTCTTA 1380
QY 1381 AGCTTTTATTTTTCCTCAACTTTTCAAAAGTCACTTCAAGAGGATATATTTTACATA 1440
DB 1381 AGCTTTTATTTTTCCTCAACTTTTCAAAAGTCACTTCAAGAGGATATATTTTACATA 1440
QY 1441 ATAAATGCACTCATTT 1458
DB 1441 ATAAATGCACTCATTT 1458

RESULT 3
US-09-111-470-9
; Sequence 9, Application US/09111470
; Patent No. 6277959
; GENERAL INFORMATION:
; APPLICANT: Valladeau, Jenny
; APPLICANT: Ravel, Odile
; APPLICANT: Bates, Elizabeth E.M.
; APPLICANT: Ford, John
; APPLICANT: Saeland, Sem
; APPLICANT: Lebecque, Serge J.E.
; TITLE OF INVENTION: Mammalian Membrane Protein Genes;
; TITLE OF INVENTION: Related Reagents
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DNAX Research Institute
; STREET: 901 California Avenue


```

; APPLICANT: Saeland, Sem
; TITLE OF INVENTION: Isolated Mammalian Membrane Protein Genes; Related Reagents
; FILE REFERENCE: SF0695 B
; CURRENT APPLICATION NUMBER: US/09/862,802A
; CURRENT FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/053,080
; PRIOR FILING DATE: 1997-07-09
; PRIOR APPLICATION NUMBER: US 09/111,470
; PRIOR FILING DATE: 1998-07-08
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9
; LENGTH: 1370
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: mammalian nucleic acid and protein
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (273)..(1091)
; OTHER INFORMATION: protein coding sequence
; US-09-862-802A-9

```

[illegible]

```

RESULT 5
US-09-949-016-4091
; Sequence 4091, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 4091
; LENGTH: 1224
; TYPE: DNA

```

ORGANISM: Human
US-09-949-016-4091

Query Match		26.2%;	Score 382;	DB 4;	Length 1224;
Best Local Similarity		66.1%;	Pred. No. 1.3e-92;		
Matches 646;		Conservative	0;	Mismatches 245;	Indels 87; Gaps 3;
QY	224	CTGTGCAACCTCAGAGCCGCTGTTGGCCCAAGCATGACAAAGGACGCTATGAAACTTCCAG	283		
DB	140	CTGAGCAATCCAGTCCAGCGCCATCATGACCAAGAGTATCAGACCTTCAG	199		
QY	284	TACTTGGAGAAATAGGTGAAGTCCAG---GGGTTTAAATAATGGGCCACTTCTCTCCAG	340		
DB	200	CATCTGGACAATGAGGAGAGTACCACCATCAGCTCAGAAAGGGCCACTCTCTCCCCAG	259		
QY	341	TCCCTCTCGCAGGCTCTCGCTGCGCCCTGCGCATCTCTGCTGCTGCTGCTGCTGCTGCTG	400		
DB	260	CCCTCTCTGAGGCTCTGCTCTCGGACCTCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	319		
QY	401	CTGCTGCTGCTCATCTCTGTTGGATTCCAAATTCCAAATTCAGAGGGAC	460		
DB	320	CTCTCTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG	379		
QY	461	CTGGTGACCTTGAGAACAGATTTTAGCAACTTTCAGACCTTCCAG	520		
DB	380	CTGCGGGCTGAGAGAGAGCTTTCAGCACTTCACAGCGAGCACGGAGGCCAGAGGTCAAG	439		
QY	521	GCATGACTTCCAGGGCAGCAGTGGAGAAACGATAGCATCTCTGAAAGCTGAGGTG	580		
DB	440	GGCTTGAGCACTCCAGGAGGCAATGTTGGGAAGAAGATGAAGTCGTAGAGTCCCAAGCTG	499		
QY	581	GAGGGTTTCAAGCAGGAAACGGCAGGAGGGGTATCTGAGCTCCAGGAACACACTACGCAG	640		
DB	500	GAGAACAGCAGAGGACCTGAGTG-----	524		
QY	641	AAGGCACACTAGGCCACTGTCCCCACTGCCCATCTGTGTGTGTGTGTGTGTGTGTGTGTG	700		
DB	525	-----AAGATCACTCCAGC	538		
QY	701	ATGCTCTGAGTCCAGCAGCTGTTGCAAGACCTGAGAACTGACCTGACCTGCGAGTGGCT	760		
DB	539	CTGCTGCTCAGCTGAAGCAGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG	598		
QY	761	ACTCTCAACAAATGCTCCACTGAAGGACCTGCTGCCCGCTCAACTGGGTGGAGCAC	820		
DB	599	GCCTCCAGGGCAATGGCT---CAGAAAGACCTGCTGCCCGCTCAACTGGGTGGAGCAC	655		
QY	821	CAAGAAGCTGTACTGTGTTCTCTCACTGTGGAGTGTCTTGGGCCGAGGCTGAGAAATAC	880		
DB	656	GAGCGCAGCTGTACTGTTCTCTGCTCCGGAGGCTTGGGCTGACGCCGACAACTAC	715		
QY	881	TGCCAGCTGAAGAACGCCACCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG	940		
DB	716	TGCGCGCTGGAGACGCGCACCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG	775		
QY	941	CAGAAATATCTAGGCTCCGATACACTGTGATGGGCTCAGTGACCTCAAGAGGAGCTGG	1000		
DB	776	GAGCAACATAGCCCTGTGAACACTGTGATGGGCTTCAAGCAACAAACGGGCCCTGG	835		
QY	1001	AAGTGGGTGGAATGGAACAGACTATGCGACCGGCTTCCAGAACTGGAAGCCAGGCCACCA	1060		
DB	836	AAGTGGGTGAGCGGACGGACTACGAGACGGGCTTCAAGAACTGGAGGCGGAGCAGCGC	895		
QY	1061	GAGCACTGGCAGGGGACCGGGCTGGGTGGAGGGGAGACTGTGCTCACTTCATCCAGAC	1120		
DB	896	GAGCACTGGTACGCGCACCGGGCTCGAGGAGGGCGAGGACTGTGTCCCACTTCAACCGAC	955		
QY	1121	GGCAGGTGGAATGACAGAGCTCTGCCAGAGGGCCCTACACTGGGTCTGCGAGGCTGGCCTG	1180		
DB	956	GGCGGTGGNAACGACGAGCTCTGCCAGAGGGCCCTACCGTGGGTCTGCGAGACAGAGCTG	1015		
QY	1181	GGTCAGACCAAGCCAGGAG	1198		

DB1016GACAGGCCAGCCAGGAG1033

RESULT 6

US-09-016-434-1186

Sequence 1186, Application US/09016434

Patent No. 6500938

GENERAL INFORMATION:

APPLICANT: Janice Au-Young

APPLICANT: Jeffrey J. Seilhamer

TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING

TITLE OF INVENTION: PATHWAY GENE EXPRESSION

NUMBER OF SEQUENCES: 1490

CORRESPONDENCE ADDRESS:

ADDRESSEE: INCYTE PHARMACEUTICALS, INC.

STREET: 3174 PORTER DRIVE

CITY: PALO ALTO

STATE: CALIFORNIA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/016,434

FILING DATE: HERewith

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Zeller, Karen J.

REGISTRATION NUMBER: 37,071

REFERENCE/DOCKET NUMBER: PA-0002 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: (650) 855-0555

TELEFAX: (650) 845-4166

INFORMATION FOR SEQ ID NO: 1186:

SEQUENCE CHARACTERISTICS:

LENGTH: 1277 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: GENBANK

CLONE: g179078

US-09-016-434-1186

Query Match26.2%; Score 382; DB 4; Length 1277;

Best Local Similarity66.1%; Pred. No. 1.3e-92;

Matches 646; Conservative 0; Mismatches 245; Indels 87;

QY224CTGTGACAACTCAGAGCGCTGTTGGCCCAAGCATGACAAAGGACGCTATGAAACTTCCAG

DB140CTGAGCAATCCAGTCCAGCTCCAGCGCCATCATGACCAAGAGTATCAGACCTTCAG

QY284TACTTGGAGAAATAGGTGAAGTCCAG---GGGTTTAAATAATGGGCCACTTCTCTCCAG

DB200CATCTGGACAATGAGGAGAGTACCACCATCAGCTCAGAAAGGGCCACTCTCTCCCCAG

QY341TCCCTCTCGCAGGCTCTCGCTGCGCCCTGCGCATCTCTGCTGCTGCTGCTGCTGCTGCTG

DB260CCCTCTCTGAGGCTCTGCTCTCGGACCTCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT

QY401CTGCTGCTGCTCATCTCTGTTGGATTCCAAATTCCAAATTCAGAGGGAC

DB320CTCTCTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG

QY461CTGGTGACCTTGAGAACAGATTTTAGCAACTTTCAGACCTTCCAG

APPLICANT: Bougueleret, Lydie
APPLICANT: Chumakov, Ilya
APPLICANT: Cohen, Annick
TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
FILE REFERENCE: GEN-T114XC2D1
CURRENT APPLICATION NUMBER: US/10/170,097
CURRENT FILING DATE: 2002-06-10
PRIOR APPLICATION NUMBER: US 09/641,638
PRIOR FILING DATE: 2000-08-16
PRIOR APPLICATION NUMBER: US 09/502,330
PRIOR FILING DATE: 2000-02-11
PRIOR APPLICATION NUMBER: US 60/133,200
PRIOR FILING DATE: 1999-05-07
PRIOR APPLICATION NUMBER: US 09/275,267
PRIOR FILING DATE: 1999-03-23
PRIOR APPLICATION NUMBER: US 60/119,917
PRIOR FILING DATE: 1999-02-12
NUMBER OF SEQ ID NOS: 1304
SOFTWARE: Patent.pm
SEQ ID NO 46
LENGTH: 400
TYPE: DNA
ORGANISM: Homo Sapiens
FEATURE:
NAME/KEY: allele
LOCATION: 128
OTHER INFORMATION: 12-198-128 : polymorphic base A or G
FEATURE:
NAME/KEY: misc_binding
LOCATION: 129..148
OTHER INFORMATION: 12-198-128.mis1, potential
FEATURE:
NAME/KEY: primer_bind
LOCATION: 1..20
OTHER INFORMATION: upstream amplification primer
FEATURE:
NAME/KEY: primer_bind
LOCATION: 380..400
OTHER INFORMATION: downstream amplification primer, complement
FEATURE:
NAME/KEY: misc_binding
LOCATION: 116..140
OTHER INFORMATION: 12-198-128 potential probe
FEATURE:
NAME/KEY: misc_feature
LOCATION: 249
OTHER INFORMATION: n=a, g, c or t
US-10-170-097-46
Query Match 12.6%; Score 183; DB 4; Length 400;
Best Local Similarity 98.9%; Pred. No. 3.1e-39;
Matches 183; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Qy 1 GTTGAGGAGATGGATGTCCAGATGATAGGGCTCTGGGATTTTCAGACCCCAAGACCAGC 60
Db 8 GTTGAGGAGATGGATGTCCAGATGATAGGGCTCTGGGATTTTCAGACCCCAAGACCAGC 67
Qy 61 AGGACTCCAGTCACTTACCCAGCTCTCAGGACACAGCGCTCCCAACTCTGAGTGAC 120
Db 68 AGGACTCCAGTCACTTACCCAGCTCTCAGGACACAGCGCTCCCAACTCTGAGTGAC 127
Qy 121 GTCCCACTCTGCTCTTGAGGACCAACCAACGTGGGAATCACACCTTCAGACCTCCCA 180
Db 128 RTCCCACTCTGCTCTTGAGGACCAACCAACGTGGGAATCACACCTTCAGACCTCCCA 187
Qy 181 CAGCT 185
Db 188 CAGT 192

RESULT 11

US-09-641-638-49
Sequence 49, Application US/09641638
Patent No. 6432648
GENERAL INFORMATION:
APPLICANT: Blumenfeld, Marta
APPLICANT: Bougueleret, Lydie
APPLICANT: Chumakov, Ilya
APPLICANT: Cohen, Annick
TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
FILE REFERENCE: GENSET.051CPI
CURRENT APPLICATION NUMBER: US/09/641,638
CURRENT FILING DATE: 2000-08-16
PRIOR APPLICATION NUMBER: US 09/502,330
PRIOR FILING DATE: 2000-02-11
PRIOR APPLICATION NUMBER: US 60/133,200
PRIOR FILING DATE: 1999-05-07
PRIOR APPLICATION NUMBER: US 09/275,267
PRIOR FILING DATE: 1999-03-23
PRIOR APPLICATION NUMBER: US 60/119,917
PRIOR FILING DATE: 1999-02-12
NUMBER OF SEQ ID NOS: 1304
SOFTWARE: Patent.pm
SEQ ID NO 49
LENGTH: 448
TYPE: DNA
ORGANISM: Homo Sapiens
FEATURE:
NAME/KEY: allele
LOCATION: 129
OTHER INFORMATION: 12-214-129 : polymorphic base C or T
NAME/KEY: misc_binding
LOCATION: 110..128
OTHER INFORMATION: 12-214-129.mis1
NAME/KEY: misc_binding
LOCATION: 130..149
OTHER INFORMATION: 12-214-129.mis2, potential complement
NAME/KEY: primer_bind
LOCATION: 1..20
OTHER INFORMATION: upstream amplification primer
NAME/KEY: primer_bind
LOCATION: 429..448
OTHER INFORMATION: downstream amplification primer, complement
NAME/KEY: misc_binding
LOCATION: 117..141
OTHER INFORMATION: 12-214-129 potential probe
US-09-641-638-49
Query Match 10.7%; Score 155.6; DB 3; Length 448;
Best Local Similarity 99.4%; Pred. No. 8e-32;
Matches 155; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy 776 GCTCCACTGAAGGACCTCTGCCCGTCAACTGGGTGAGCACCAAGACAGCTGTCTAC 835
Db 103 GCCTCCACTGAAGGACCTCTGCCCGTCAACTGGGTGAGCACCAAGACAGCTGTCTAC 162
Qy 836 TGGTTCCTCTCACTCTGGGATGTCTGGCCGAGGCTGAGAAGTACTGCCAGCTGAAGAC 895
Db 163 TGGTTCCTCTCACTCTGGGATGTCTGGCCGAGGCTGAGAAGTACTGCCAGCTGAAGAC 222
Qy 896 GCCCACCCTGGTGTCTCACTCAACTCCAGGGAGGAGCAG 931
Db 223 GCCCACCCTGGTGTCTCACTCAACTCCAGGGAGGAGCAG 258

RESULT 12

US-10-170-097-49
Sequence 49, Application US/10170097
Patent No. 6794143
GENERAL INFORMATION:

```
/ APPLICANT: Blumenfeld, Marta
/ APPLICANT: Bougueleret, Lydie
/ APPLICANT: Chumakov, Ilya
/ APPLICANT: Cohen, Annick
/ TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
/ FILE REFERENCE: GEN-T114XC2D1
/ CURRENT APPLICATION NUMBER: US/10/170,097
/ CURRENT FILING DATE: 2002-06-10
/ PRIOR APPLICATION NUMBER: US 09/641,638
/ PRIOR FILING DATE: 2000-08-16
/ PRIOR APPLICATION NUMBER: US 09/502,330
/ PRIOR FILING DATE: 2000-02-11
/ PRIOR APPLICATION NUMBER: US 60/133,200
/ PRIOR FILING DATE: 1999-05-07
/ PRIOR APPLICATION NUMBER: US 09/275,267
/ PRIOR FILING DATE: 1999-03-23
/ PRIOR APPLICATION NUMBER: US 60/119,917
/ PRIOR FILING DATE: 1999-02-12
/ NUMBER OF SEQ ID NOS: 1304
/ SOFTWARE: Patent.pm
/ SEQ ID NO 49
/ LENGTH: 448
/ TYPE: DNA
/ ORGANISM: Homo Sapiens
/ FEATURE:
/ NAME/KEY: allele
/ LOCATION: 129
/ OTHER INFORMATION: 12-214-129 : polymorphic base C or T
/ FEATURE:
/ NAME/KEY: misc_binding
/ LOCATION: 110..128
/ OTHER INFORMATION: 12-214-129.mis1
/ FEATURE:
/ NAME/KEY: misc_binding
/ LOCATION: 130..149
/ OTHER INFORMATION: 12-214-129.mis2, potential complement
/ FEATURE:
/ NAME/KEY: primer_bind
/ LOCATION: 1..20
/ OTHER INFORMATION: upstream amplification primer
/ FEATURE:
/ NAME/KEY: primer_bind
/ LOCATION: 429..448
/ OTHER INFORMATION: downstream amplification primer, complement
/ FEATURE:
/ NAME/KEY: misc_binding
/ LOCATION: 117..141
/ OTHER INFORMATION: 12-214-129 potential probe
/ US-10-170-097-49

Query Match          10.7%; Score 155.6; DB 4; Length 448;
Best Local Similarity 99.4%; Pred. No. 8e-32; Indels 0; Gaps 0;
Matches 155; Conservative 1; Mismatches 0;

QY 776 GCCTCCACTGAAGGACCTCTGCTGCCCGTCAACTGGGTGGAGCACCAGACAGCTGCTAC 835
DB 103 GCCTCCACTGAAGGACCTCTGCCCYGTCAACTGGGTGGAGCACCAGACAGCTGCTAC 162
QY 836 TGGTTCTCTCACTCTGGATGTCTGGGCCGAGGCTGAGAAGTACTGCCAGCTGAAGAAC 895
DB 163 TGGTTCTCTCACTCTGGATGTCTGGGCCGAGGCTGAGAAGTACTGCCAGCTGAAGAAC 222
QY 896 GCCCACTGTGTGTATCACTCACTCCAGGAGGAGCAG 931
DB 223 GCCCACTGTGTGTATCACTCCAGGAGGAGCAG 258

RESULT 13
US-09-641-638-51
; Sequence 51, Application US/09641638
; Patent No. 6432648
; GENERAL INFORMATION:
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Chumakov, Ilya
; APPLICANT: Cohen, Annick
; TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
; FILE REFERENCE: GENSET.051CP1
```

```
/ APPLICANT: Blumenfeld, Marta
/ APPLICANT: Bougueleret, Lydie
/ APPLICANT: Chumakov, Ilya
/ APPLICANT: Cohen, Annick
/ TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
/ FILE REFERENCE: GENSET.051CP1
/ CURRENT APPLICATION NUMBER: US/09/641,638
/ CURRENT FILING DATE: 2000-08-16
/ PRIOR APPLICATION NUMBER: US 09/502,330
/ PRIOR FILING DATE: 2000-02-11
/ PRIOR APPLICATION NUMBER: US 60/133,200
/ PRIOR FILING DATE: 1999-05-07
/ PRIOR APPLICATION NUMBER: US 09/275,267
/ PRIOR FILING DATE: 1999-03-23
/ PRIOR APPLICATION NUMBER: US 60/119,917
/ PRIOR FILING DATE: 1999-02-12
/ NUMBER OF SEQ ID NOS: 1304
/ SOFTWARE: Patent.pm
/ SEQ ID NO 51
/ LENGTH: 448
/ TYPE: DNA
/ ORGANISM: Homo Sapiens
/ FEATURE:
/ NAME/KEY: allele
/ LOCATION: 358
/ OTHER INFORMATION: 12-214-360 : polymorphic base G or C
/ NAME/KEY: misc_binding
/ LOCATION: 338..357
/ OTHER INFORMATION: 12-214-360.mis1, potential
/ NAME/KEY: misc_binding
/ LOCATION: 359..378
/ OTHER INFORMATION: 12-214-360.mis2, potential complement
/ NAME/KEY: primer_bind
/ LOCATION: 1..20
/ OTHER INFORMATION: upstream amplification primer
/ NAME/KEY: primer_bind
/ LOCATION: 429..448
/ OTHER INFORMATION: downstream amplification primer, complement
/ NAME/KEY: misc_binding
/ LOCATION: 346..370
/ OTHER INFORMATION: 12-214-360 potential probe
/ US-09-641-638-51

Query Match          10.6%; Score 154.4; DB 3; Length 448;
Best Local Similarity 99.4%; Pred. No. 1.7e-31;
Matches 155; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 776 GCCTCCACTGAAGGACCTGTGCTGCCCGTCAACTGGGTGGAGCACCAGACAGCTGCTAC 835
DB 103 GCCTCCACTGAAGGACCTGTGCTGCCCTGTCAACTGGGTGGAGCACCAGACAGCTGCTAC 162
QY 836 TGGTTCTCTCACTCTGGATGTCTGGGCCGAGGCTGAGAAGTACTGCCAGCTGAAGAAC 895
DB 163 TGGTTCTCTCACTCTGGATGTCTGGGCCGAGGCTGAGAAGTACTGCCAGCTGAAGAAC 222
QY 896 GCCCACTGTGTGTATCACTCACTCCAGGAGGAGCAG 931
DB 223 GCCCACTGTGTGTATCACTCCAGGAGGAGCAG 258

RESULT 14
US-09-641-638-571
; Sequence 571, Application US/09641638
; Patent No. 6432648
; GENERAL INFORMATION:
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Chumakov, Ilya
; APPLICANT: Cohen, Annick
; TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
; FILE REFERENCE: GENSET.051CP1
```


; CURRENT APPLICATION NUMBER: US/09/641,638
; CURRENT FILING DATE: 2000-08-16
; PRIOR APPLICATION NUMBER: US 09/502,330
; PRIOR FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: US 60/133,200
; PRIOR FILING DATE: 1999-05-07
; PRIOR APPLICATION NUMBER: US 09/275,267
; PRIOR FILING DATE: 1999-03-23
; PRIOR APPLICATION NUMBER: US 60/119,917
; PRIOR FILING DATE: 1999-02-12
; NUMBER OF SEQ ID NOS: 1304
; SOFTWARE: Patent.pm
; SEQ ID NO 571
; LENGTH: 448
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: allele
; LOCATION: 85
; OTHER INFORMATION: 12-214-85 : deletion CCTAT
; NAME/KEY: misc_binding
; LOCATION: 65..84
; OTHER INFORMATION: 12-214-85.misl, potential
; NAME/KEY: primer_bind
; LOCATION: 1..20
; OTHER INFORMATION: upstream amplification primer
; NAME/KEY: primer_bind
; LOCATION: 429..448
; OTHER INFORMATION: downstream amplification primer, complement
; OTHER INFORMATION: 12-214-85 : deletion CCTAT

Query Match 10.6%; Score 154.4; DB 3; Length 448;
Best Local Similarity 99.4%; Pred. No. 1.7e-31;
Matches 155; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy	776	GCCTCCACTGAAGGGACCTCTGCCCGTCAACTGGGTGGAGCACCAGACAGCTGCTAC	835
Db	103	GCCTCCACTGAAGGGACCTCTGCCCGTCAACTGGGTGGAGCACCAGACAGCTGCTAC	162
Qy	836	TGTTTCTCTCACTCTGGGATGTCTGGGCGAGGCTGAGAAGTACTGCCACGCTGAAGAAC	895
Db	163	TGTTTCTCTCACTCTGGGATGTCTGGGCGAGGCTGAGAAGTACTGCCACGCTGAAGAAC	222
Qy	896	GCCACCTGGTGTCTATCACTCACTCCAGGAGGAGCAG	931
Db	223	GCCACCTGGTGTCTATCACTCACTCCAGGAGGAGCAG	258

RESULT 15

US-10-170-097-51
; Sequence 51, Application US/10170097
; Patent No. 6794143
; GENERAL INFORMATION:
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Chumakov, Ilya
; APPLICANT: Cohen, Annick
; TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
; FILE REFERENCE: GEN-T114XC2D1
; CURRENT APPLICATION NUMBER: US/10/170,097
; CURRENT FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: US 09/641,638
; PRIOR FILING DATE: 2000-08-16
; PRIOR APPLICATION NUMBER: US 09/502,330
; PRIOR FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: US 60/133,200
; PRIOR FILING DATE: 1999-05-07
; PRIOR APPLICATION NUMBER: US 09/275,267
; PRIOR FILING DATE: 1999-03-23
; PRIOR APPLICATION NUMBER: US 60/119,917
; PRIOR FILING DATE: 1999-02-12
; NUMBER OF SEQ ID NOS: 1304

; SOFTWARE: Patent.pm
; SEQ ID NO 51
; LENGTH: 448
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: allele
; LOCATION: 358
; OTHER INFORMATION: 12-214-360 : polymorphic base G or C
; FEATURE:
; NAME/KEY: misc_binding
; LOCATION: 338..357
; OTHER INFORMATION: 12-214-360.misl, potential
; FEATURE:
; NAME/KEY: misc_binding
; LOCATION: 359..378
; OTHER INFORMATION: 12-214-360.misl, potential complement
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 1..20
; OTHER INFORMATION: upstream amplification primer
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 429..448
; OTHER INFORMATION: downstream amplification primer, complement
; FEATURE:
; NAME/KEY: misc_binding
; LOCATION: 346..370
; OTHER INFORMATION: 12-214-360 potential probe
; OTHER INFORMATION: 12-214-360 potential probe
US-10-170-097-51

Query Match 10.6%; Score 154.4; DB 4; Length 448;
Best Local Similarity 99.4%; Pred. No. 1.7e-31;
Matches 155; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy	776	GCCTCCACTGAAGGGACCTCTGCCCGTCAACTGGGTGGAGCACCAGACAGCTGCTAC	835
Db	103	GCCTCCACTGAAGGGACCTCTGCCCGTCAACTGGGTGGAGCACCAGACAGCTGCTAC	162
Qy	836	TGTTTCTCTCACTCTGGGATGTCTGGGCGAGGCTGAGAAGTACTGCCACGCTGAAGAAC	895
Db	163	TGTTTCTCTCACTCTGGGATGTCTGGGCGAGGCTGAGAAGTACTGCCACGCTGAAGAAC	222
Qy	896	GCCACCTGGTGTCTATCACTCACTCCAGGAGGAGCAG	931
Db	223	GCCACCTGGTGTCTATCACTCACTCCAGGAGGAGCAG	258

Search completed: May 27, 2005, 23:57:16

Job time : 271 secs

This Page Blank (uspto)

Result No.	Query			ID	Description		
	Score	Match	Length				
C	1	1277.2	87.6	1610	3	CRG20226	full-leng
	2	841.6	57.7	1134	5	BQ067335	AGENCOURT
	3	759.4	52.1	816	5	BQ018186	UI-H-DP1-
	4	736	40.5	827	1	ALS46075	ALS46075
	5	709.6	48.7	842	4	BI820020	603035394
C	6	706.2	48.4	829	4	BI821633	603036669
	7	658.4	45.2	678	5	BUG21880	UI-H-DF0-
C	8	656	45.0	674	6	CA412958	UI-H-E20-
C	9	655.8	45.0	721	4	BI838214	BI838214
	10	654.8	44.9	710	5	BUG687211	UI-CF-ECL
	11	654	44.9	774	4	BI820664	603034446
C	12	649	44.5	728	4	BI838868	603087264
	13	637	43.7	663	6	CA418829	UI-H-E21-
	14	621.4	42.6	634	4	BI905917	BI905917
	15	616.6	42.3	816	4	BI167532	603061425
	16	604.6	41.5	852	4	BI910159	603069529
C	17	604	41.4	671	4	BM714409	UI-E-EJO-
	18	592.6	40.6	935	4	BI912448	603291140
	19	592.2	40.6	720	4	BI910082	603067920
	20	582.4	39.9	824	4	BI838052	603086617
	21	578	39.6	721	4	BI518752	603061934
C	22	567.4	38.9	718	4	BI910338	603068394
	23	565.8	38.8	583	5	BP228397	BP228397
	24	563.8	38.7	600	5	BUG16457	UI-H-DF0-

QY	65	CTCCAGTCACTTACCCAGCTCTCCAGGACACAGCGTCCCAACTCTGAGTGAGCTCC	124
Db	61	CTCCAGTCACTTACCCAGCTCTCCAGGACACAGCGTCCCAACTCTGAGTGAGCTCC	120
QY	125	CACCTCTGGTCTTGAGACAAACAAAGTGGGAATCACACCTCCAGACCTCCCAACAGC	184
Db	121	CACCTCTGGTCTTGAGACAAACAAAGTGGGAATCACACCTCCAGACCTCCCAACAGC	180
QY	185	TCACACCCAGACTGGCGCGCCCTGCGCTCCATTTTCAGTGTGACAACTCAGAGCGCT	244
Db	181	TCACACCCAGACTGGCGCGCCCTGCGCTCCATTTTCAGTGTGACAACTCAGAGCGCT	240
QY	245	GTGGGCCCAAGCATGACAGGACGTATGAAAACTTCAGTACTTTGAGAGAAATAGGTGAAA	304
Db	241	GTGGGCCCAAGCATGACAGGACGTATGAAAACTTCAGTACTTTGAGAGAAATAGGTGAAA	300
QY	305	GTCCAGGGGTTAAAAATAGGCGCACCTTCCTCTCCAGTCCCTCTGAGCGGTCTCCGCTCT	364
Db	301	GTCCAGGGGTTAAAAATAGGCGCACCTTCCTCTCCAGTCCCTCTGAGCGGTCTCTGCTCT	360
QY	365	GGGCGCTGCATCTCTGCTGCTCCCTGGGCGCTGGGCTGCTGCTGCTGCTCATCTCTCT	424
Db	361	GGGCGCTGCATCTCTGCTGCTCCCTGGGCGCTGGGCTGCTGCTGCTGCTCATCTCTCT	420
QY	425	GTGGTTGGATTCCAAATTCAGAGGAGCTGTGTGACCTCTGAGAACAGATTTT	484
Db	421	GTGGTTGGATTCCAAATTCAGAGGAGCTGTGTGACCTCTGAGAACAGATTTT	480
QY	485	AGCAACTTCACCTCAAACTGTGGCGGAGATCCAGGCACTGACTTCCAGGGCAGCAGC	544
Db	481	AGCAACTTCACCTCAAACTGTGGCGGAGATCCAGGCACTGACTTCCAGGGCAGCAGC	540
QY	545	TTGGAGAAACGATACATCTCTGAAGCTGAGTGGAGGTTTCAAGCAGGAACGGCAG	604
Db	541	TTGGAGAAACGATACATCTCTGAAGCTGAGTGGAGGTTTCAAGCAGGAACGGCAG	600
QY	605	GCAGGGGTATCTGAGCTCCAGGAACACACTACGACAGAGCACACCTAGGCCACTGTCCC	664
Db	601	G-----	601
QY	665	CACGTGCCCATCTGTGTGTCTGAGTTCATTTGAAATGCTCTGCGAGTCCAGCAGCTG	724
Db	602	-----CAGTTTCATTTCTGAAATGCTCTGCGAGTCCAGCAGCTG	639
QY	725	GTGCAAGACTGAAGAACTGACCTGCCAGGTGGCTACTCTCAACCAATGCTCCACT	784
Db	640	GTGCAAGACTGAAGAACTGACCTGCCAGGTGGCTACTCTCAACCAATGCTCCACT	699
QY	785	GAAGGACCTGTGCTGCCCGCTCAACTGGGTGGAGCACCAAGACAGCTGCTACTGTTCTCT	844
Db	700	GAAGGACCTGTGCTGCCCGCTGTCACTGGGTGGAGCACCAAGACAGCTGCTACTGTTCTCT	759
QY	845	CACCTCTGGATGTCTGGGCGAGGTGAGAACTATGTCAGGTGAAGAACGCCACCTG	904
Db	760	CACCTCTGGATGTCTGGGCGAGGTGAGAACTATGTCAGGTGAAGAACGCCACCTG	819
QY	905	GTGGTTCATCAACTCCAGGGAGGACAGAAATTTTGTCCAGAAATATCTAGGCTCCGCATAC	964
Db	820	GTGGTTCATCAACTCCAGGGAGGACAGAAATTTTGTCCAGAAATATCTAGGCTCCGCATAC	879
QY	965	ACCTGGATGGGCTCTAGTCACTCCAGGAGCTTGAAGTGGGTGGATGGAAACAGACTAT	1024
Db	880	ACCTGGATGGGCTCTAGTCACTCCAGGAGCTTGAAGTGGGTGGATGGAAACAGACTAT	939
QY	1025	GCACCGGCTTCCAGAACTGGAAAGCCAGGCGACGCCAGACGATGTCGGAGGGGACAGGGCTG	1084
Db	940	GCACCGGCTTCCAGAACTGGAAAGCCAGGCGACGCCAGACGATGTCGGAGGGGACAGGGCTG	999
QY	1085	GGTGGAGGCGGACACTGTGCTCACTTCCATCCAGAGCGGAGGTGGAAATGACGACGCTGTC	1144
Db	1000	GGTGGAGGCGGAGGACTGTGCTCACTTCCATCCAGAGCGGAGGTGGAAATGACGACGCTGTC	1059

QY	1145	CAGAGCCCTACCACTGGGTCTGCGAGGCTGGCTGGGTGAGAGGAGAGTCTAC	1204
Db	1060	CAGAGCCCTACCACTGGGTCTGCGAGGCTGGCTGGGTGAGAGGAGAGTCTAC	1119
QY	1205	TCAGCTGCTTTGGTGGGACACCCGGCCACAGAAATGGCGGTGGAGGAGGACTCTTCT	1264
Db	1120	TCAGCTGCTTTGGTGGGACACCCGGCCACAGAAATGGCGGTGGAGGAGGACTCTTCT	1179
QY	1265	CACGACCTCTCGCAAGACCGCTCTGGGAGAGAAATAAGCACTGGGAGATTGGAAGCACT	1324
Db	1180	CACGACCTCTCGCAAGACCGCTCTGGGAGAGAAATAAGCACTGGGAGATTGGAAGCACT	1239
QY	1325	GCTAACATTTTGAATTTTTTCTCTTAATTTTAAAAAGATGCTAGTGTCTTTAAGCT	1384
Db	1240	GCTAACATTTTGAATTTTTTCTCTTAATTTTAAAAAGATGCTAGTGTCTTTAAGCT	1299
QY	1385	TTTATTTTTTTTCCAACTTTTGAAGTCACTTCAATGAAGGTATTAATTTTTTACATAATA	1444
Db	1300	TTTATTTTTTTTCCAACTTTTGAAGTCACTTCAATGAAGGTATTAATTTTTTACATAATA	1359
QY	1445	AAATGCACATTTT 1458	
Db	1360	AAATGCACATTTT 1373	
RESULT 2			
BO067335			
LOCUS			
DEFINITION			
5', mRNA sequence.			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
FEATURES			
source			
1. .1134			
/organism="Homo sapiens"			
/mol_type="mRNA"			
/db_xref="taxon:9606"			
/clone="IMAGE:5751735"			
/lab_host="DH10B"			
/clone_lib="NIH_MGC_115"			
/note="Organ: pooled brain, lung, testis; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: EcoRV (destroyed); RNA source anonymous Pool of 6 male brains, age range 23-27; 1 male lung, age 27; and 1 male testis, age 69. Library is oligo-dr primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.8 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 021. Note: this is a NIH_MGC Library."			

ORIGIN

Query Match 57.7%; Score 841.6; DB 5; Length 1134;

QY	806	AACTGGGTGGAGCACAAGACAGCTGCTACTTGTTCTCTCACTCTGGGATGCTCTGGGCC	865
Db	671	AACTGGGTGGAGCACCAGACAGCTGCTACTTGTTCTCTCACTCTGGGATGCTCTGGGCC	612
QY	866	GAGGCTGAGAGTACTGCCAGCTGAAGAGCCACCCTGCTGTTGTCATCAACTCCAGGGAG	925
Db	611	GAGGCTGAGAGTACTGCCAGCTGAAGAGCCACCCTGCTGTTGTCATCAACTCCAGGGAG	552
QY	926	GAGCAAAATTTTCTCCAGAAATATCTAGGCTCGCATACACCTTGGATGGGCTCAGTGAC	985
Db	551	GAGCAAAATTTTCTCCAGAAATATCTAGGCTCGCATACACCTTGGATGGGCTCAGTGAC	492
QY	986	CCTGAAGAGCCCTGGAGTGGGTGGATGGAACAGACTATGCGACCGGCTTCAGAACTGG	1045
Db	491	CCTGAAGAGCCCTGGAGTGGGTGGATGGAACAGACTATGCGACCGGCTTCAGAACTGG	432
QY	1046	AAGCCAGGCCAGCAGACACTGGCAGGGGACCGGCTGGGTGGAGCGAGACTGTGCT	1105
Db	431	AAGCCAGGCCAGCAGACACTGGCAGGGGACCGGCTGGGTGGAGCGAGACTGTGCT	372
QY	1106	CACCTCCATCCAGACGGCAGGTGGAATGACGCTCTGCCAGAGGCCCTACCACCTGGGTC	1165
Db	371	CACCTCCATCCAGACGGCAGGTGGAATGACGCTCTGCCAGAGGCCCTACCACCTGGGTC	312
QY	1166	TGCGAGGCTGGCTGGGTGAGACCCAGGAGAGTCACTGAGCTGGCTTTTGGTGGACC	1225
Db	311	TGCGAGGCTGGCTGGGTGAGACCCAGGAGAGTCACTGAGCTGGCTTTTGGTGGACC	252
QY	1226	ACCGGCCACAGAAATGGCGGTGGAGAGGACTCTTCTCAGACCTCTCGAAGACCG	1285
Db	251	ACCGGCCACAGAAATGGCGGTGGAGAGGACTCTTCTCAGACCTCTCGAAGACCG	192
QY	1286	CTCTGGGAGAGAAATAGCACCTGGGAGATTGGAAGCACCTGCTAAATTTTGAATTTTTT	1345
Db	191	CTCTGGGAGAGAAATAGCACCTGGGAGATTGGAAGCACCTGCTAAATTTTGAATTTTTT	132
QY	1346	CTCTTTAAATTTTAAAGATGGTATAGTGTCTTTAAGCTTTTATTTTTTTTCCAACTTTT	1405
Db	131	CTCTTTAAATTTTAAAGATGGTATAGTGTCTTTAAGCTTTTATTTTTTTTCCAACTTTT	72
QY	1406	GAAAGTCACTTCATGAGGTATAATTTTATCAATAATAAAATGCACTCATTT	1458
Db	71	GAAAGTCACTTCATGAGGTATAATTTTATCAATAATAAAATGCACTCATTT	19
RESULT 4			
AL546075			
LOCUS			
DEFINITION			
AL546075 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA			
clone CS0D1024YN10 5-PRIME, mRNA sequence.			
ACCESSION			
AL546075.3 GI:45746544			
VERSION			
EST.			
SOURCE			
Homo sapiens (human)			
ORGANISM			
Homo sapiens			
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
1 (bases 1 to 827)			
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.			
Full-length cDNA libraries and normalization			
Unpublished (2001)			
On Feb 15, 2001 this sequence version replaced gi:31267909.			
Contact: Genoscope			
Genoscope - Centre National de Sequencage			
2 rue Gaston Cremlieux, CP 5706 - 91057 EVRY cedex - FRANCE			
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr			
1st strand cDNA was primed with a NotI-oligo (dT) primer. Five prime			
end enriched, double-strand cDNA was digested with Not I and cloned			
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library			
was normalized. Library was constructed by life technologies, a			
division of invitrogen. This sequence belongs to sequence cluster			
8440.r			
For more information about this cluster, see			

FEATURES		http://www.genoscope.cns.fr/cdna?8=CS0D1024G05QP1&c=8440.r.	
source		Location/Qualifiers	
		1..827	
		/organism="Homo sapiens"	
		/mol_type="mRNA"	
		/db_xref="taxon:9606"	
		/clones="CS0D1024YN10"	
		/tissue type="PLACENTA COT 25-NORMALIZED"	
		/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"	
		/note="1st strand cDNA was primed with a NotI-oligo (dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."	
ORIGIN		Query Match 50.5%; Score 736; DB 1; Length 827;	
		Best Local Similarity 91.1%; Pred. No. 8.6e-181;	
		Matches 827; Conservative 0; Mismatches 0; Indels 81; Gaps 1;	
QY	5	AGGAGATGGGATGTCCAGATGATAGGGCTCCTGGGATTTTCAGACCCCAAGACCAGCAGGA	64
Db	1	AGGAGATGGGATGTCCAGATGATAGGGCTCCTGGGATTTTCAGACCCCAAGACCAGCAGGA	60
QY	65	CTCCAGTCACTCTACCCAGCTCTCCAGGACACAGCGCTCCCAACTCTGAGTGAAGTCC	124
Db	61	CTCCAGTCACTCTACCCAGCTCTCCAGGACACAGCGCTCCCAACTCTGAGTGAAGTCC	120
QY	125	CACCTCTGGTCTTTCAGACACCAACAGTGGGAATCACACCTCCAGACCTCCCAAGC	184
Db	121	CACCTCTGGTCTTTCAGACACCAACAGTGGGAATCACACCTCCAGACCTCCCAAGC	180
QY	185	TCCACCCAGACTGGGGCGCGGCTGCTCCATTTTCAGCTGTGACAACTTCAGAGCCGT	244
Db	181	TCCACCCAGACTGGGGCGCGGCTGCTCCATTTTCAGCTGTGACAACTTCAGAGCCGT	240
QY	245	GTTGGCCCAAGCATGACAAGGACGTATGAAAACTTCCAGTACTTTGGAGAAATAGGTGAA	304
Db	241	GTTGGCCCAAGCATGACAAGGACGTATGAAAACTTCCAGTACTTTGGAGAAATAGGTGAA	300
QY	305	GTCCAGGGGTTTAAAAATGGGGCCACTTCTCCAGTCCCTCCAGCGCTCCGCTCT	364
Db	301	GTCCAGGGGTTTAAAAATGGGGCCACTTCTCCAGTCCCTCCAGCGCTCCGCTCT	360
QY	365	GGGGCTCCCATCTCTGCTGCTGGGCTCGGCTCGGCTGCTGCTGCTCATCATCTGT	424
Db	361	GGGGCTCCCATCTCTGCTGCTGGGCTCGGCTCGGCTGCTGCTGCTCATCATCTGT	420
QY	425	GTGGTTGGATTTCCAAAAATTCAAAAATTTTCAGAGGACCTGGTGAACCTGAGAACAGATTTT	484
Db	421	GTGGTTGGATTTCCAAAAATTCAAAAATTTTCAGAGGACCTGGTGAACCTGAGAACAGATTTT	480
QY	485	AGCACTTTCACCTCAACACACTGTGGCGGAGATCCAGGCACTGACTTCCAGGGGACGAGC	544
Db	481	AGCACTTTCACCTCAACACACTGTGGCGGAGATCCAGGCACTGACTTCCAGGGGACGAGC	540
QY	545	TTGGAAGAACGATAGCATCTCTGAAAGCTGAGGTGGAGGGTTTCAAGCAGGAAACGGCAG	604
Db	541	TTGGAAGAACGATAGCATCTCTGAAAGCTGAGGTGGAGGGTTTCAAGCAGGAAACGGCAG	600
QY	605	GCAGGGGTATCTGAGCTCCAGGAACACACTACGCAAGAGGCAACCTTAGGCGCACTGTCCC	664
Db	601	G-----	601
QY	665	CACCTCCCATCTGTGTGTGCTCCAGTTCAATTCGAAATGCTCTGCGAGTCCAGCAGCTG	724
Db	602	-----	639
QY	725	GTGCAAGACCTGAAGAAATGACCTGCGCAGGTGGCTACTCTCAACAACTAGCTCCACT	784
Db	640	GTGCAAGACCTGAAGAAATGACCTGCGCAGGTGGCTACTCTCAACAACTAGCTCCACT	699
QY	785	GAAGGACCTGCTGCGGCTCAACTGGGTGGAGCAACCAAGACAGCTGTCTGCTGTTCT	844


```

Db      700 GAAGGACCTGCTGCCCGTCAACTGGGTGGAGCACAAGACAGCTGCTACTGTTCTCT 759
Qy      845 CACTCTGGGATGTCCTGGGCGGAGGCTGAGAAGTACTGCGAGCTGAAGAACGCCCACTG 904
Db      760 CACTCTGGGATGTCCTGGGCGGAGGCTGAGAAGTACTGCGAGCTGAAGAACGCCCACTG 819
Qy      905 GTGGTCTAT 912
Db      820 GTGGTCTAT 827

RESULT 5
BI820020
LOCUS      842 bp mRNA linear EST 04-OCT-2001
DEFINITION 60303534F1 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5176487 5',
mRNA sequence.
ACCESSION BI820020
VERSION BI820020.1 GI:15931570
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 842)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11439 Row: 1 Column: 24
High quality sequence stop: 819.
Location/Qualifiers
1. 842
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5176487"
/lab_host="DH10B"
/clone_lib="NIH MGC 115"
/notes="Organ: pooled brain, lung, testis; Vector:
pCMV-SPORT6; Site 1: NotI; Site 2: EcoRV (destroyed); RNA
source anonymous pool of 6 male brains, age range 23-27; 1
male lung, age 27; and 1 male testis, age 69. Library is
oligo-dT primed and directionally cloned (EcoRV site is
destroyed upon cloning). Average insert size 1.8 kb,
insert size range 1-3 kb. Library is normalized and
enriched for full-length clones and was constructed by C.
Gruber (Invitrogen). Research Genetics tracking code
021. Note: this is a NIH_MGC Library."

FEATURES
source
Query Match 48.7%; Score 709.6; DB 4; Length 842;
Best Local Similarity 90.5%; Pred. No. 6.9e-174;
Matches 836; Conservative 0; Mismatches 4; Indels 84; Gaps 4;

Qy      1 GTTGAGGAGATGGGATGTCCTCCAGATGATAGGGCTCTCTGGGATTTCCAGACCAAGACGAGC 60
Db      1 GTTGAGGAGATGGGATGTCCTCCAGATGATAGGGCTCTCTGGGATTTCCAGACCAAGACGAGC 60
Qy      61 AGGACTCCAGTCACTTACCCAGGCTCTCCAGGACACAGCGGCTCCCAACTCTGAGTGAC 120
Db      1 GTTGAGGAGATGGGATGTCCTCCAGATGATAGGGCTCTCTGGGATTTCCAGACCAAGACGAGC 60
Qy      61 AGGACTCCAGTCACTTACCCAGGCTCTCCAGGACACAGCGGCTCCCAACTCTGAGTGAC 120
Db      1 GTTGAGGAGATGGGATGTCCTCCAGATGATAGGGCTCTCTGGGATTTCCAGACCAAGACGAGC 60
Qy      121 GTCCCACTCTGCTCTTGGAGCAGCACCAACGTTGGGAATCACACCTCCAGACCTCCCA 180
Db      121 GTCCCACTCTGCTCTTGGAGCAGCACCAACGTTGGGAATCACACCTCCAGACCTCCCA 180

```

```

Qy      181 CAGCTCAACCCAGAGACTGGGCGCGGCGCTCCATTTTCAGCTGTGACAACTCTAGAG 240
Db      181 CAGCTCAACCCAGAGACTGGGCGCGGCGCTCCATTTTCAGCTGTGACAACTCTAGAG 240
Qy      241 CCGTGTGGCCCAAGCATGACAGGACGATATGAAAATTTCCAGTACTTGGAGATAAGGT 300
Db      241 CCGTGTGGCCCAAGCATGACAGGACGATATGAAAATTTCCAGTACTTGGAGATAAGGT 300
Qy      301 GAAAGTCCAGGGGTTTAAAAATGGGCCACTTCTCTCCAGTCCCTCTCTGAGGCTCTCCG 360
Db      301 GAAAGTCCAGGGGTTTAAAAATGGGCCACTTCTCTCCAGTCCCTCTCTGAGGCTCTCTG 360
Qy      361 CTCTGGGCGCTGCGCATCTCTCTGCTGCTCCCTGGGCGCTGGGCTGCTGCTGCTCATCAT 420
Db      361 CTCTGGGCGCTGCGCATCTCTCTGCTGCTCCCTGGGCGCTGGGCTGCTGCTGCTCATCAT 420
Qy      421 CTGCTGTGTTGGATTCCAAAATTTCCAAAATTTCCAGAGGACCTCTGTAACCTTGAGAACAGA 480
Db      421 CTGCTGTGTTGGATTCCAAAATTTCCAAAATTTCCAGAGGACCTCTGTAACCTTGAGAACAGA 480
Qy      481 TTTTAGCAACTTCACTCAAACTCTGGCGGAGATCCAGGCACTGACTTTCCAGGGCAG 540
Db      481 TTTTAGCAACTTCACTCAAACTCTGGCGGAGATCCAGGCACTGACTTTCCAGGGCAG 540
Qy      541 CAGCTTGGAGAAACGATAGCATCTCTGAAAGCTGAGGTGGAGGGTTTCAAGCAGGAAACG 600
Db      541 CAGCTTGGAGAAACGATAGCATCTCTGAAAGCTGAGGTGGAGGGTTTCAAGCAGGAAACG 600
Qy      601 GCAGGACGAGGGTATCTGAGCTCCAGGAACACACTACGCAAGAGGCACACCTAGGCCACTG 660
Db      601 GCAGG-----
Qy      661 TCCCCACTGCCCCACTCTGTGTGTGCCAGTTCATTTCTGAAATGCTCTCTGAGAGTCCAGCA 720
Db      606 -----CAGTTCAATCTGAAATGCTCTCTGAGAGTCCAGCA 639
Qy      721 GCTGTGTGAAGACCTGGAAGAACTGACCTGCGCAGGTGGCTACTCTCTCAACAAATGCTCTC 780
Db      640 GCTGTGTGAAGACCTGGAAGAACTGACCTGCGCAGGTGGCTACTCTCTCAACAAATGCTCTC 699
Qy      781 CACTGAAGGACCTGCTGCCCCGCTCAACTGGGTGGAGACCAAGACAGCTGCTACTGTT 840
Db      700 CACTGAAGGACCTGCTGCTG-----CCCTCAACTGGGTGGAGACCAAGACAGCTGCTACTGTT 758
Qy      841 CTCTCACTCT-GGGATGTCCTGGGCGGAGGCTGAGAACTGAGAGTACTGCCAGCTGAAGAACGCC 899
Db      759 CTCTCACTCTGGGGATGTCCTGGGCGGAGGCTGAGAACTGAGAGTACTTGCAGCTGAAGAACGCC 818
Qy      900 ACCT-GGTGTCTATCAACTCCAGG 922
Db      819 ACCTGGGTGTCTATCAACTCCAGG 842

RESULT 6
BI821633
LOCUS
DEFINITION 60303669F1 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5177584 5',
mRNA sequence.
ACCESSION BI821633
VERSION BI821633.1 GI:15933183
KEYWORDS EST.
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 829)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.

```

cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 cDNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLAM11442 row: j column: 17
 High quality sequence stop: 823.

FEATURES

```

Location/Qualifiers
1..829
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5177584"
/lab_host="DH10B"
/clone_lib="NIH MGC 115"
/notes="Organ: pooled brain, lung, testis; Vector:
pCMV-SPORT6; Site 1: NotI; Site 2: EcoRV (destroyed); RNA
source anonymous pool of 6 male brains, age range 23-27; 1
male lung, age 27; and 1 male testis, age 69. Library is
oligo-dT primed and directionally cloned (EcoRV site is
destroyed upon cloning). Average insert size 1.8 kb,
insert size range 1-3 kb. Library is normalized and
enriched for full-length clones and was constructed by C.
Gruber (Invitrogen). Research Genetics tracking code
021. Note: this is a NIH_MGC Library."

ORIGIN
Query Match 48.4%; Score 706.2; DB 4; Length 829;
Best Local Similarity 90.5%; Pred. No. 5.3e-173;
Matches 821; Conservative 0; Mismatches 3; Indels 83; Gaps 3;

QY 10 ATGGGATGTCACGATGATAGGCTCTCGGATTTTCAGACCCCAAGACAGGAGTCCCA 69
DB 1 ATGGGATGTCACGATGATAGGCTCTCGGATTTTCAGACCCCAAGACAGGAGTCCCA 60

QY 70 GTCACCTTACCCAGCTCTCCAGGACACAGCGCTCCCAACTCTGAGTGACGTCCCACT 129
DB 61 GTCACCTTACCCAGCTCTCCAGGACACAGCGCTCCCAACTCTGAGTGACGTCCCACT 120

QY 130 CTGGTCTTTCAGCACAAACAGTGGGAATCACACCTCCAGACTCCACAGCTCCAC 189
DB 121 CTGGTCTTTCAGCACAAACAGTGGGAATCACACCTCCAGACTCCACAGCTCCAC 180

QY 190 CCAGACTGGGCGCGGCTGCTCCATTTAGCTGTGACCACTTCAGACCGGTGTGG 249
DB 181 CCAGACTGGGCGCGGCTGCTCCATTTAGCTGTGACCACTTCAGACCGGTGTGG 240

QY 250 CCCAAGCATGACAAAGACGTATGAAACTTCCAGTACTTTGGAGATAAGGTGAAAGTCCA 309
DB 241 CCCAAGCATGACAAAGACGTATGAAACTTCCAGTACTTTGGAGATAAGGTGAAAGTCCA 300

QY 310 GGGTTTAAATATGGGCGACTTCTCTCAGTCCCTCTCGAGCGTCTCGCTCTGGGCC 369
DB 301 GGGTTTAAATATGGGCGACTTCTCTCAGTCCCTCTCGAGCGTCTCGCTCTGGGCC 360

QY 370 CTGCCATCTCTCTCTCGGCGCTCGGCGCTGCTGCTGCTGCTCATCATCTGTGTGT 429
DB 361 CTGCCATCTCTCTCTCGGCGCTCGGCGCTGCTGCTGCTGCTCATCATCTGTGTGT 420

QY 430 TGGATTCCAAATATCCAAATTTTCAGAGGACCTGTGACCCCTTGAGAACAGATTTTAGCAA 489
DB 421 TGGATTCCAAATATCCAAATTTTCAGAGGACCTGTGACCCCTTGAGAACAGATTTTAGCAA 480

QY 490 CTTACCTCAACACTGTGGCGAGATCCAGGCACTGACTTCCAGGCGAGCAGCTTGA 549
DB 481 CTTACCTCAACACTGTGGCGAGATCCAGGCACTGACTTCCAGGCGAGCAGCTTGA 540

QY 550 AGAAACGATAGCATCTCTGAAAGCTCAGGTGGAGGTTTCAAGCAGGAACCGCAGG 609
DB 541 AGAAACGATAGCATCTCTGAAAGCTCAGGTGGAGGTTTCAAGCAGGAACCGCAGG 600

QY 610 GGTATCTGAGTCCAGGAACACACTAGCAGGAAGGCACACCTAGGCGCACTGTCCCACTG 669

```

```

Db 597 ----- 596
QY 670 CCCATCTGTGTGCCAGTTTCATTCTGAATGCTCTCGAGTCCAGAGCTGTGCA 729
Db 597 -----CAGTTTCATTCTGAATGCTCTCGAGTCCAGAGCTGTGCA 639
QY 730 AGACCTGAAGAAACTGACCTGCCAGGTGGTCTCTCAACCAACAACTGCTCCACTGAAGG 789
Db 640 AGACCTGAAGAAACTGACCTGCCAGGTGGTCTCTCAACCAACAACTGCTCCACTGAAGG 699
QY 790 GACCTGTGCCCCGTCAACTGGGTGGAGACCAAGACAGCTGCTACTGTGTTCTCTCACTC 849
Db 700 GACCTGTGCCCCGTCAACTGGGTGGAGCACCAAGACAGCTGCTACTGTGTTCTCTCACTC 759
QY 850 TGGGATGTCCTGGGCGCGAGG-CTGAGAAGTACTGCCAGCTGAAGACGCCACCTGTGG 908
Db 760 TGGGATGTCCTGGGCGCGAGGCTGAGAAGTACTGCCAGCTGAAGACG-CCACCTGTGG 818
QY 909 TCATCAA 915
Db 819 TCATCAA 825

RESULT 7
LOCUS BU621880 678 bp mRNA linear EST 23-SEP-2002
DEFINITION UI-H-DF0-bev-j-20-0-UI.s1 NCI CGAP_DF0 Homo sapiens cDNA clone
ACCESSION UI-H-DF0-bev-j-20-0-UI 3', mRNA sequence.
VERSION BU621880
KEYWORDS BU621880.1 GI:23288095
SOURCE EST.
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 678)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. Jose Mercuende
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be obtained
from Dr. M. Bento Soares, bento-soares@uiowa.edu
The following repetitive elements were found in this cDNA
sequence: 1-22, >AT rich#Low_complexity (matched complement)
Seq primer: M13 FORWARD
POLYA=Yes.
Location/Qualifiers
1..678
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-H-DF0-bev-j-20-0-UI"
/tissue_type="Subchondral Bone"
/dev_stage="Adult"
/clone_lib="NCI CGAP DF0"
/notes="Organ: Bone; Vector: pVT73-Pac (Pharmacia) with a
modified polylinker; Site 1: EcoR I; Site 2: Not I;
NCI CGAP DF0 is a cDNA library containing the following
tissues: Subchondral Bone. The library was constructed
according to Bonaldo, Lennon and Soares, Genome Research,
6:791-806, 1996. First strand cDNA synthesis was primed
with an oligo-dT primer containing a Not I site. Double
stranded cDNA was ligated to an EcoR I adaptor, digested
with Not I, and cloned directionally into pVT73-Pac
vector. The oligonucleotide used to prime the synthesis of

```

first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is GTTAAGCGTC.
TAG TISSUE=chondrosarcoma bone
TAG LIB=UI-H-DF0
TAG_SEQ=GTTAAGCGTC

ORIGIN

Query Match 45.2%; Score 658.4; DB 5; Length 678;
Best Local Similarity 99.8%; Pred. No. 1.6e-160;
Matches 659; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 799 CCCGCTCAACTGGGTGGAGCACCAGACAGCTGCTACTGTTCTCTCACTCTGGATGTC 858
DB 678 CCTGTCACTGGGTGGAGCACCAGACAGCTGCTACTGTTCTCTCACTCTGGATGTC 619
QY 859 CTGGCCGAGGCTGAGAGTACTGCCAGCTGAGAACGCCACCTCGTGGTGCATCAACTC 918
DB 618 CTGGCCGAGGCTGAGAGTACTGCCAGCTGAGAACGCCACCTCGTGGTGCATCAACTC 559
QY 919 CAGGAGGAGCAGAAATTTTGTCCAGAAATATCTAGGCTCCGCATACACCTGGATGGGCCT 978
DB 558 CAGGAGGAGCAGAAATTTTGTCCAGAAATATCTAGGCTCCGCATACACCTGGATGGGCCT 499
QY 979 CAGTGACCTGAGAGGCTGGAAGTGGGTGGATGGACAGACTATCGACCGCTTCCA 1038
DB 498 CAGTGACCTGGAAGGAGCCTGGAGTGGGTGGATGGACAGACTATCGACCGCTTCCA 439
QY 1039 GAACTGGAAGCCAGGCCAGCAGACGCTGGCAGGGGCA CGGGCTGGGTGGAGCGAGGA 1098
DB 438 GAACTGGAAGCCAGGCCAGCAGACGCTGGCAGGGGCA CGGGCTGGGTGGAGCGAGGA 379
QY 1099 CTGTGCTCACTTCAATCCAGACGGCAGGTGGAATGACGACGTCTGCCAGAGGCCCTACCA 1158
DB 378 CTGTGCTCACTTCAATCCAGACGGCAGGTGGAATGACGACGTCTGCCAGAGGCCCTACCA 319
QY 1159 CTGGGTCTGAGGCTGGCTGGGTGACAGCAGCCAGAGAGTCACTGAGTGGCTTTGG 1218
DB 318 CTGGGTCTGAGGCTGGCTGGGTGACAGCAGCCAGGAGTCACTGAGTGGCTTTGG 259
QY 1219 TGGGACCAACCGGCCACAGAAATGGCGGTGGAGGAGGACTCTTCTCAAGACCTCTCTGC 1278
DB 258 TGGGACCAACCGGCCACAGAAATGGCGGTGGAGGAGGACTCTTCTCAAGACCTCTCTGC 199
QY 1279 AAGACCGCTCTGGGAGAGAAATAGCACTGGAGATTGGAAGCACTCTCAACATTTGAA 1338
DB 198 AAGACCGCTCTGGGAGAGAAATAGCACTGGAGATTGGAAGCACTCTCAACATTTGAA 139
QY 1339 TTTTCTCTTTAAATTTTAAAGATGGTATAGTGTCTTAAGCTTTTATTTTTC 1398
DB 138 TTTTCTCTTTAAATTTTAAAGATGGTATAGTGTCTTAAGCTTTTATTTTTC 79
QY 1399 AACTTTGAAAGTCACTTCAATGAAGTATATTTTACATAATAAATAATGCACTCAATT 1458
DB 78 AACTTTGAAAGTCACTTCAATGAAGTATATTTTACATAATAAATAATGCACTCAATT 19

RESULT 8
CA412958/c

LOCUS CA412958 674 bp mRNA linear EST 07-NOV-2002
DEFINITION UI-H-E20-bao-b-08-0-UI.s1 NCI_CGAP_Ch1 Homo sapiens cDNA clone
UI-H-E20-bao-b-08-0-UI 3', mRNA sequence.

ACCESSION

VERSION CA412958

KEYWORDS CA412958.1 GI:24775609

SOURCE EST.

ORGANISM Homo sapiens (human)

Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 674)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP).

Tumor Gene Index

JOURNAL
COMMENT

Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-femail.nih.gov
Tissue Procurement: Dr. Steven Gitelis/ Rush Presbyterian, Dept. of
Orthopaedics
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be obtained
from Dr. M. Bento Soares, bento-soares@uiowa.edu
The following repetitive elements were found in this cDNA
sequence: 1-22, >AT rich#Low_complexity (matched complement)
Seq primer: M13 FORWARD
POLYA-Yes.

FEATURES

source

1. .674
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-H-E20-bao-b-08-0-UI"
/tissue_type="Chondrosarcoma Grade II"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies)"
/clone_lib="NCI CGAP Ch1"
/note="Organ: Left Pelvis; Vector: pT7T3-Pac (Pharmacia)
with a modified polylinker; Site_1: Ecor I; Site_2: Not I;
NCI CGAP Ch1 is a cDNA library containing the following
tissue(s): Chondrosarcoma Grade II. The library was
constructed according to Bonaldo, Lennon and Soares,
Genome Research, 6:791-806, 1996. First strand cDNA
synthesis was primed with an oligo-dT primer containing a
Not I site. Double stranded cDNA was ligated to an Ecor I
adaptor, digested with Not I, and cloned directionally
into pT7T3-Pac vector. The oligonucleotide used to prime
the synthesis of first-strand cDNA contains a library tag
sequence that is located between the Not I site and the
(dT)18 tail. The sequence tag for this library is
TGATCAGCT.
TAG TISSUE=grade-2-chondrosarcoma
TAG LIB=UI-H-E20
TAG_SEQ=ATCTAATATG

ORIGIN

Query Match 45.0%; Score 656; DB 6; Length 674;
Best Local Similarity 100.0%; Pred. No. 6.6e-160;
Matches 656; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 803 GTCAACTGGGTGGAGCACCAGACAGCTGCTACTGGTTCCTCACTCTGGGATGTCCTGG 862
DB 674 GTCAACTGGGTGGAGCACCAGACAGCTGCTACTGGTTCCTCACTCTGGGATGTCCTGG 615
QY 863 GCCGAGGCTGAGAAGTACTGCCAGCTGGAAGAGCCACCTGGTGGTGCATCAACTCCAGG 922
DB 614 GCCGAGGCTGAGAAGTACTGCCAGCTGGAAGAGCCACCTGGTGGTGCATCAACTCCAGG 555
QY 923 GAGGAGCAGAAATTTTGTCCAGAAATATCTAGGCTCCGCATACACCTGGATGGGCTCAGT 982
DB 554 GAGGAGCAGAAATTTTGTCCAGAAATATCTAGGCTCCGCATACACCTGGATGGGCTCAGT 495
QY 983 GACCTTGAAGGAGCCTGGAGTGGGTGGATGGAGACAGACTATGCGACCGCTCCAGAAC 1042
DB 494 GACCTTGAAGGAGCCTGGAGTGGGTGGATGGAGACAGACTATGCGACCGCTCCAGAAC 435
QY 1043 TGGAGCCAGGCCAGCCAGACGACTGGCAGGGGCA CGGGCTGGGTGGAGGCGAGGACTGT 1102
DB 434 TGGAGCCAGGCCAGCCAGACGACTGGCAGGGGCA CGGGCTGGGTGGAGGCGAGGACTGT 375
QY 1103 GCTCACTTCCATCCAGACGCGAGGTGGAATGACGACTCTGCCAGAGGCCCTACCACTGG 1162
DB 374 GCTCACTTCCATCCAGACGCGAGGTGGAATGACGACTCTGCCAGAGGCCCTACCACTGG 315
QY 1163 GTCTGCGAGGCTGGCTGGGTGCAGACCAGCAGGAGAGTCACTGAGTGCCTTTGGTGGG 1222

Db	314	GTCTCGAGGCTGGCCCTGGGTGAGACCAAGGAGAGTCACTGAGCTGCCCTTTGGTGGG	255
Qy	1223	ACCACCCGGCCACAGAAATGGCGGTGGGAGGAGGACTCTTCTCACGACCTCTCGCAAGA	1282
Db	254	ACCACCCGGCCACAGAAATGGCGGTGGGAGGAGGACTCTTCTCACGACCTCTCGCAAGA	195
Qy	1283	CGCTCTGGGAGAGAAATAGCACTGGGAGATTGGAAGCAGCTGCTAAACATTTTGAATTTT	1342
Db	194	CGCTCTGGGAGAGAAATAGCACTGGGAGATTGGAAGCAGCTGCTAAACATTTTGAATTTT	135
Qy	1343	TTTCTCTTAATTTTAAAGATGGTATAGTGTCTTAAGCTTTTATTTTTCCACT	1402
Db	134	TTTCTCTTAATTTTAAAGATGGTATAGTGTCTTAAGCTTTTATTTTTCCACT	75
Qy	1403	TTTGAAGTCAACTTCATGAAGGTATAATTTTACATAATAAAATGCACTCATTT	1458
Db	74	TTTGAAGTCAACTTCATGAAGGTATAATTTTACATAATAAAATGCACTCATTT	19
RESULT 9			
BI838214			
LOCUS			
DEFINITION			
603083132F1 NIH_MGC_120 Homo sapiens cDNA clone IMAGE:522622 5',			
mRNA sequence.			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
Contact: Robert Strausberg, Ph.D.			
Email: cgabbs-remail.nih.gov			
Tissue Procurement: Life Technologies, Inc.			
cDNA Library Preparation: Life Technologies, Inc.			
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)			
DNA Sequencing by: Incyte Genomics, Inc.			
Clone distribution: MGC clone distribution information can be			
found through the I.M.A.G.E. Consortium/LLNL at:			
http://image.llnl.gov			
Plate: L1AM11559 row: 0 column: 07			
High quality sequence stop: 719.			
FEATURES			
source			
1. .721			
/organism="Homo sapiens"			
/mol_type="mRNA"			
/db_xref="taxon:9606"			
/clone="IMAGE:522622"			
/lab_host="DH10B"			
/clone_lib="NIH MGC 120"			
/notes="Organ: pooled pancreas and spleen; Vector:			
pCMV-SPORT6; Site 1: NotI; Site 2: EcoRV (destroyed); RNA			
source anonymous pool of spleen and pancreas from 28 yo			
male Library is oligo-dT primed and directionally cloned			
(EcoRV site is destroyed upon cloning). Average insert			
size 1.5 kb, insert size range 1-2.5 kb. Library is			
normalized and enriched for full-length clones and was			
constructed by C. Gruber (Invitrogen). Research Genetics			
tracking code 025. Note: this is a NIH_MGC Library."			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
668; Conservative			
1; Gaps			
1; Indels			
1; Mismatches			
0; P			
0; Pred. No. 7.6e-160;			
45.0%; Score 655.8; DB 4; Length 721;			
1 GTTGAGGAGTGGGATGTCAGATAGGCTCTGGGATTTACAGCCAGACGACG			
60			
Db	41	GTTGAGGAGTGGGATGTCAGATAGGCTCTGGGATTTACAGCCAGACGACG	100

Qy	61	AGGACTCCAGTCACTCTACCCAGCTCTCCAGGACACAGCGCTCCCACTCTGAGTGAC	120
Db	101	AGGACTCCAGTCACTCTACCCAGCTCTCCAGGACACAGCGCTCCCACTCTGAGTGAC	160
Qy	121	GTCCACCTCTGGTCTTGGAGCACAAACAGTGGGAATCACACCTTCCAGACCTCCCA	180
Db	161	GTCCACCTCTGGTCTTGGAGCACAAACAGTGGGAATCACACCTTCCAGACCTCCCA	220
Qy	181	CAGCTTCCACCCAGACTGGCGCGCGCTCTCCATTTTCAGCTGTGCAACACCTTCAGAG	240
Db	221	CAGCTTCCACCCAGACTGGCGCGCGCTCTCCATTTTCAGCTGTGCAACACCTTCAGAG	280
Qy	241	CCGTGTTGGCCCAAGCATGACAAGGACGTATGAAACCTTCAGTACTTTGGGAATAAGGT	300
Db	281	CCGTGTTGGCCCAAGCATGACAAGGACGTATGAAACCTTCAGTACTTTGGGAATAAGGT	340
Qy	301	GAAGTCCAGGGGTTTAAAAATGGCCACTTCTCTCCAGTCCCTCCCTGCGGCTCTCG	360
Db	341	GAAGTCCA - GGGTTTAAAAATGGCCACTTCTCTCCAGTCCCTCCCTGCGGCTCTCG	399
Qy	361	CTCTGGGCGCTGCCATCTCTCTGCTGTCTCGGCGCTCGGCTCTGCTGTGCTGGTTCATCAT	420
Db	400	CTCTGGGCGCTGCCATCTCTCTGCTGTCTCGGCGCTCGGCTCTGCTGTGCTGGTTCATCAT	459
Qy	421	CTGTGTGTTGGATTCCAAAATTTCCAAAATTTCCAGAGGACCTGTGTGACCTCGAGAACAGA	480
Db	460	CTGTGTGTTGGATTCCAAAATTTCCAAAATTTCCAGAGGACCTGTGTGACCTCGAGAACAGA	519
Qy	481	TTTATGCAACTTCACTCAAAACACTGTGGCGGAGATCCAGGCACTGACTTCCAGGGGAG	540
Db	520	TTTATGCAACTTCACTCAAAACACTGTGGCGGAGATCCAGGCACTGACTTCCAGGGGAG	579
Qy	541	CAGCTTTGGAAGAAACGATAGCATCTCTGAAAGCTGAGGTGGAGGGTTTCAAGACGAAACG	600
Db	580	CAGCTTTGGAAGAAACGATAGCATCTCTGAAAGCTGAGGTGGAGGGTTTCAAGACGAAACG	639
Qy	601	GCAGGACGGGTATCTGAGCTCCAGGAAACACACTACGAGGAGGACACCTAGGCCACTG	660
Db	640	GCAGGACGGGTATCTGAGCTCCAGGAAACACACTACGAGGAGGACACCTAGGCCACTG	699
Qy	661	TCCCCACTGCCC	671
Db	700	TCCCCACTGCCC	710
RESULT 10			
EUG87211/c			
LOCUS			
DEFINITION			
UI-CF-EC1-adv-g-09-0-UI.s1 UI-CF-EC1 Homo sapiens cDNA clone			
UI-CF-EC1-adv-g-09-0-UI 3', mRNA sequence.			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
COMMENT			
Contact: McCray, PB			
University of Iowa			
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA			
Tel: 319 356 4866			
Fax: 319 356 7171			
Email: paul-mccray@uiowa.edu			
Tissue Procurement: Dr. M. J. Welsh, University of Iowa			
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa			

cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 cDNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research
 Genetics (www.resgen.com) or from Open Biosystems
 (www.openbiosystems.com).
 Seq primer: M13 FORWARD
 POLYA=Yes.

FEATURES

Source

1. 710
/location="Guarrilreis"
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-CF-ECL-adv-g-09-0-UI"
/tissue_type="Lung"
/dev_stage="Adult and Fetal"
/lab_host="DH10B (Life Technologies) (T1 phase resistant)"
/clone_lib="UI-CF-ECL1"
/note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site_1: Ecor I; Site 2: Not I;
UI-CF-ECL1 is a normalized cDNA library containing the following tissue(s): Normal lung from adult and from fetal day 64, day 87, week 19 and week 42. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an Ecor I Not I site. Double stranded cDNA was primed to contain a adaptor digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dr)18 tail. The sequence tag for this library is AAGTGGCTTAC.
TAG_TISSUE=Normal Lung Epithelial Cells Tissue nos 369-371 and 380-383
TAG_LIB=UI-CF-ECL1
TAG_SEQ=AAGTGGCTTAC"

ORIGIN

Query Match	44.9%;	Score	654.8;	DB	5;	Length	710.0;
Best Local Similarity	99.3%;	Pred. No.	1.4e-159;				
Matches	667;	Conservative	0;	Mismatches	4;	Indels	1; Gaps 1;
QY	787	AGGAGCTGCTGCCCGTCAACTGGGTGGAGCACCAAGACAGCTGCTACTGGTTCTCTCA	846				
DB	709	AAGGACCTGCTGCCCTGTCAACTGGGT-GAGCACCAAGACAGCTGCTACTGGTTCTCTCA	651				
QY	847	CTCTGGGATGCTCTGGSCCGAGGCTGAGAAGTACTGCGCAGCTGAAGAAACCCACACCTGGT	906				
DB	650	CTCTGNGATGCTCTGNNCCGAGGCTGAGAAGTACTGCGCAGCTGAAGAAACCCACACCTGGT	591				
QY	907	GGTCATCAACTCCAGGAGGAGCAGAAATTTTGTCCAGAAATATCTAGGCTCCGCATACAC	966				
DB	590	GGTCATCAACTCCAGGAGGAGCAGAAATTTGTCCAGAAATATCTAGGCTCCGCATACAC	531				
QY	967	CTGGATGGGCTCAGTGACCCCTGAAGAGGCTTGGAGTGGGTGGATGGAAACAGACTATGC	1026				
DB	530	CTGGATGGGCTCAGTGACCCCTGAAGAGGCTTGGAGTGGGTGGATGGAAACAGACTATGC	471				
QY	1027	GACCGGCTTCCAGAACTGGGAAGCAGCCAGCCAGCAGACGATGGCAGGGGACACGGGCTGGG	1086				
DB	470	GACCGGCTTCCAGAACTGGGAAGCAGCCAGCCAGCAGACGACTGGCAGGGGACACGGGCTGGG	411				
QY	1087	TGGAGCGAGGACTGTGCTCACTTCCATCCAGACGGCAGGTGGAAATGACACAGCTGTGCCA	1146				
DB	410	TGGAGCGAGGACTGTGCTCACTTCCATCCAGACGGCAGGTGGAAATGACACAGCTGTGCCA	351				
QY	1147	GAGGCCCTACCACTTGGGTCTGCGAGGCTGGCCTGGGTTCAGACACGCGAGGAGTCACTG	1206				
DB	350	GAGGCCCTACCACTTGGGTCTGCGAGGCTGGCCTGGGTTCAGACACGCGAGGAGTCACTG	291				
QY	1207	AGTGTGCTTTGGTGGGACCAACCCGGGCCACAGAAATGCGGTGGGAGGAGGACTCTTCTCA	1266				
DB	290	AGTGTGCTTTGGTGGGACCAACCCGGGCCACAGAAATGCGGTGGGAGGAGGACTCTTCTCA	231				

Qy	1267	CGAGCTCTCGAAGACCGCTCTGGGAGAGAAATAGCACTGGGAGATTGGGAAGCACTGC	1326
Db	230	CGAGCTCTCGAAGACCGCTCTGGGAGAGAAATAGCACTGGGAGATTGGGAAGCACTGC	171
Qy	1327	TAACTTTTCAAATTTTTTTCTTTTAATTTTAAAAAGATCGTATAGTGTCTTTAAAGCTTT	1386
Db	170	TAACTTTTGAATTTTTTTCTTTTAATTTTAAAAAGATCGTATAGTGTCTTTAAAGCTTT	111
Qy	1387	TATTTTTTTTCCAACTTTTGAAGAGTCAACTTCATGAAGGTATATATTTTACATAATAAAA	1446
Db	110	TATTTTTTTTCCAACTTTTGAAGAGTCAACTTCATGAAGGTATATATTTTACATAATAAAA	51
Qy	1447	ATGCACTCATTT 1458	
Db	50	ATGCACTCATTT 39	
RESULT 11			
BI820664			
LOCUS	BI820664	774 bp mRNA linear EST 04-OCT-2001	
DEFINITION	603034446F1 NIH MGC_115 Homo sapiens cDNA clone IMAGE:5175683 5'		
ACCESSION	BI820664	mRNA sequence.	
VERSION	BI820664.1	GI:15932214	
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 774)		
TITLE	NIH-MGC http://mgc.nci.nih.gov/ .		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		
COMMENT	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgabs-r@mail.nih.gov		
	Tissue Procurement: Life Technologies, Inc.		
	cDNA Library Preparation: Life Technologies, Inc.		
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)		
	DNA sequencing by: Incyte Genomics, Inc.		
	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at:		

ORIGIN

	Query Match	44.9%	Score 654;	DB 4;	Length 774;
	Best Local Similarity	96.5%	Pred. No. 2.3e-159;		
	Matches 693;	Conservative	0;	Mismatches 15;	Indels 10; Gaps 2;
Qy	750	GCCAGGTGGCTACTCTCAACAAACAAAT-----GCCTCCACTGAAGGACCTGTGCTGCC	800		
Db	1	GCCAGGTGGCTACTCTCAACAAACAAATGGTGAGGAAGCCTCACTGAAGGACCTGTGCTGCC	60		

```
QY 801 CCGTCAACTGGGTGGAGCACCAGACAGCTGCTACTGTTCTCTCACTCTGGGATGTCT 860
Db 61 CCGTCAACTGGGTGGAGCACCAGACAGCTGCTACTGTTCTCTCACTCTGGGATGTCT 120
QY 861 GGGCCGAGGCTGAGAAGTACTGCCAGCTGAAGAACGCCACCTGTTGTTCACTCACTCCA 920
Db 121 GGGCCGAGGCTGAGAAGTACTGCCAGCTGAAGAACGCCACCTGTTGTTCACTCACTCCA 180
QY 921 GGGAGGAGCAGAAATTTGTCCAGAAATATCTAGGCTCCGCATACACCTCGATGGGCTCA 980
Db 181 GGGAGGAGCAGAAATTTGTCCAGAAATATCTAGGCTCCGCATACACCTCGATGGGCTCA 240
QY 981 GTGACCTCAAGGAGGCTCGAAGTGGGTGGATGGAAACAGACTATGCGACCGCTTCCAGA 1040
Db 241 GTGACCTCAAGGAGGCTCGAAGTGGGTGGATGGAAACAGACTATGCGACCGCTTCCAGA 300
QY 1041 ACTGGAAGCCAGCCAGCCAGACGACTGCGAGGGGACCGGGCTGGGTGGAGCGGAGCT 1100
Db 301 ACTGGAAGCCAGCCAGCCAGACGACTGCGAGGGGACCGGGCTGGGTGGAGCGGAGCT 360
QY 1101 GTGCTCACTTCCATCCAGACGGCAGCTGGAATGACGAGCTCTGCCAGAGGCGCTTACCCT 1160
Db 361 GTGCTCACTTCCATCCAGACGGCAGCTGGAATGACGAGCTCTGCCAGAGGCGCTTACCCT 420
QY 1161 GGGTCTGCGAGGCTGGCCTGGGTGAGCCAGAGGAGTCACTGAGCTGCTTTGGTG 1220
Db 421 GGGTCTGCGAGGCTGGCCTGGGTGAGCCAGAGGAGTCACTGAGCTGCTTTGGTG 480
QY 1221 GGACACCCGGCCACAGAAATGGCGTGGGAGGAGGACTCTTCTCAGACCTCTCCGCA 1280
Db 481 GGACACCCGGCCACAGAAATGGCGTGGGAGGAGGACTCTTCTCAGACCTCTCCGCA 540
QY 1281 GACCGCTCTGGGAGAGAAATAGCACTGGGAGATTGGAAGCACTGTAACATTTTGAATT 1340
Db 541 GACCGCTCTGGGAGAGAAATAGCACTGGGAGATTGGAAGCACTGTAACATTTTGAATT 600
QY 1341 TTTTCTCTTTAAATTTTAAAGATGGTATAGTGTCTTTAAGCTTTTATTTTTTCCAA 1400
Db 601 TTTTCTCTTTAAATTTTAAAGATGGTATAGTGTCTTTAAGCTTTTATTTTTTCCAA 660
QY 1401 CTTTGA-AGTCACTTCATGAGGTATATATTTTACATATATATAATGCACTCAAT 1457
Db 661 CTTTGAAGTCACTTCATGAGGTATATATTTTACATATATATAATGCACTCAAT 718

RESULT 12
BI83888
LOCUS 603087264F1 NIH_MGC_120 Homo sapiens cDNA clone IMAGE:5226471 5',
DEFINITION mRNA sequence.
ACCESSION BI83888.1 GI:15950438
VERSION EST.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE
AUTHORS Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
JOURNAL 1 (bases 1 to 728)
COMMENT NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L14M11569 row: o column: 16
High quality sequence start: 7
High quality sequence stop: 728.
```

```
FEATURES
source
Location/Qualifiers
1..728
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5226471"
/lab_host="DH10B"
/clone_lib="NIH_MGC_120"
/note="Organ: pooled pancreas and spleen; Vector:
pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
source anonymous pool of spleen and pancreas from 28 yo
male. Library is oligo-dT primed and directionally cloned
(EcoRV site is destroyed upon cloning). Average insert
size 1.5 kb, insert size range 1-2.5 kb. Library is
normalized and enriched for full-length clones and was
constructed by C. Gruber (Invitrogen), Research Genetics
tracking code 025. Note: this is a NIH_MGC Library."
ORIGIN
Query Match 44.5%; Score 649; DB 4; Length 728;
Best Local Similarity 99.4%; Pred. No. 4.5e-158;
Matches 682; Conservative 0; Mismatches 1; Indels 3; Gaps 3;
QY 3 TGAGAGATGGGATGTCCAGATGATAGGGGCTCTCTGGGATTTCCAGACCAGACGAGCAG 62
Db 45 TGAGAGATGGGATGTCCAGATGATAGGGGCTCTCTGGGATTTCCAGACCAGACGAGCAG 104
QY 63 GACTCCAGTCACTTACCCAGCTCTCCAGGACACAGCGCTCCCACTCTGAGTGAGCT 122
Db 105 GACTCCAGTCACTTACCCAGCTCTCCAGGACACAGCGCTCCCACTCTGAGTGAGCT 164
QY 123 CCCACCTCTGCTCTT-GCAGCACAAACCAACGTTGGGAATCACACCTCCAGACCTCCAC 181
Db 165 CCCACCTCTGCTCTTGGCAGCACAAACGTTGGGATCACACCTCCAGACCTCCAC 224
QY 182 AGCTCCACCCAGACTGGGCGCCCTGCTCTCATTTTCACTGTGTGACAACTCCAGAGC 241
Db 225 AGCTCCACCCAGACTGGGCGCCCTGCTCTCATTTTCACTGTGTGACAACTCCAGAGC 284
QY 242 CGTGTGGCCCAACGATGACAGGACGATGATAAACTTCCAGTACTTGGAGAAATAGGTG 301
Db 285 CGTGTGGCCCAACGATGACAGGACGATGATAAACTTCCAGTACTTGGAGAAATAGGTG 344
QY 302 AAAGTCCAGGGGTTTAAATAATGGGCCACTTCTCTCCAGTCCCTCTCCAGCGTCTCCGC 361
Db 345 AAAGTCCA-GGGTTTAAATAATGGGCCACTTCTCTCCAGTCCCTCTCCAGCGTCTCCGC 403
QY 362 TCTGGGCGCTGCCATCTCTCTGCTGCTGGGCGCTCGGCGCTGCTGCTGCTGCTCATATC 421
Db 404 TCTGGGCGCTGCCATCTCTCTGCTGCTGGGCGCTCGGCGCTGCTGCTGCTCATATC 463
QY 422 TGTGTGGTTGGATTTCAAAATTTCCAAATTTCCAGAGGGACCTGGTGACCTGAGACAGAT 481
Db 464 TGTGTGGTTGGATTTCAAAATTTCCAAATTTCCAGAGGGACCTGGTGAGACCTGAGACAGAT 523
QY 482 TTTAGCAACTTTCACCTCAAAACACTGTGGCGGAGATCCAGGCACTGACTTCCAGGGCAGC 541
Db 524 TTTAGCAACTTTCACCTCAAAACACTGTGGCGGAGATCCAGGCACTGACTTCCAGGCGCAGC 583
QY 542 AGCTTTGGAAGAAACGATAGCATCTCTGAAAGCTGAGGTGGAGGGTTTCAAGCAGGAAACGG 601
Db 584 AGCTTTGGAAGAAACGATAGCATCTCTGAAAGCTGAGGTGGAGGGTTTCAAGCAGGAAACGG 643
QY 602 CAGGCGAGGGTATCTGAGCTCCAGGACACACTAGGAGAGGCGACACTAGGCGCAGCTGT 661
Db 644 CAGGCGAGGGTATCTGAGCTCCAGGAAACACACTAGGAGAGGCGACACTAGGCGCAGCTGT 703
QY 662 CCCCACTGCCCATCTGTGTGTGTGCC 687
Db 704 CCCCACTG-CCATCTGTGTGTGTGCC 728
```

RESULT 13

CA418829/c
LOCUS CA418829 663 bp mRNA linear EST 07-NOV-2002
DEFINITION UI-H-EZ1-bbg-m-10-0-UI.s1 NCI CGAP Ch2 Homo sapiens cDNA clone
ACCESSION UI-H-EZ1-bbg-m-10-0-UI 3', mRNA sequence.
VERSION CA418829
KEYWORDS EST.
SOURCE CA418829.1 GI:24781480
ORGANISM Homo sapiens (human)
REFERENCE 1 (bases 1 to 663)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-@email.nih.gov
Tissue Procurement: Dr. Steven Gitelis/ Rush Presbiterian, Dept. of
Orthopedics
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be obtained
from Dr. M. Bento Soares, bento-soares@uiowa.edu
Seq primer: M13 FORWARD
POLYA=Yes.

FEATURES
source
1..663
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-H-EZ1-bbg-m-10-0-UI"
/tissue_type="Chondrosarcoma Grade II"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies)"
/clone_lib="NCI CGAP Ch2"
/notes="Organ: Left Pelvis; Vector: pT73-Pac (Pharmacia)
with a modified polylinker; Site 1: Ecor I; Site 2: Not I;
NCI CGAP Ch2 is a normalized cDNA library containing the
following tissue(s): Chondrosarcoma Grade II. The library
was constructed according to Bonaldo, Lennon and Soares,
Genome Research, 6:791-806, 1996. First strand cDNA
synthesis was primed with an oligo-dT primer containing a
Not I site. Double stranded cDNA was ligated to an Ecor I
adaptor, digested with Not I, and cloned directionally
into pT73-Pac vector. The oligonucleotide used to prime
the synthesis of first-strand cDNA contains a library tag
sequence that is located between the Not I site and the
(dT)18 tail. The sequence tag for this library is
TGATCACGCT.
TAG_TISSUE=grade-2-chondrosarcoma
TAG_LIB=UI-H-EZ1
TAG_SEQ=ATCTAATG"

ORIGIN
Query Match 43.7%; Score 637; DB 6; Length 663;
Best Local Similarity 99.8%; Pred. No. 6.1e-155;
Matches 637; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 821 CAAGACAGCTGCTACTGGTCTCTCACTCGGGATGTCCTGGCGCGAGGCTGAGAGTAC 880
DB 663 CAAGACAGCTGCTACTGGTCTCTCACTCGGGATGTCCTGGCGCGAGGCTGAGAGTAC 604
QY 881 TGCAGCTGAAGAACGCCACCTGGTGTATCACTCAAGTCCAGGAGGAGGAGCAATTTTGTG 940
DB 603 TGCAGCTGAAGAACGCCACCTGGTGTATCACTCAAGTCCAGGAGGAGGAGCAATTTTGTG 544
QY 941 CAGAAATATCTAGGCTCCGCATACACCTGGATGGGCTCTAGTGACCTCGAAGGAGCCTGG 1000
DB 543 CAGAAATATCTAGGCTCCGCATACACCTGGATGGGCTCTAGTGACCTCGAAGGAGCCTGG 484
QY 1001 AAGTGGGTGGATGGACAGACTATGCGACCGGGCTTCAGAACTGGAAAGCCAGGCCAGCCA 1060

483 AAGTGGGTGGATGGACAGACTATGCGACCGGCTTCAGAACTGGAAAGCCAGGCCAGCCA 424
QY 1061 GACGACTGGCAGGGGCAACGGGCTGGGTGGAGGCGAGGAGCTGTGCTCACTTCCATCCAGAC 1120
DB 423 GACGACTGGCAGGGGCAACGGGCTGGGTGGAGGCGAGGAGCTGTGCTCACTTCCATCCAGAC 364
QY 1121 GGCAGGTGGAAATGACGACGCTCTGCCAGAGGCCCTACACTGGGTCTGCGAGGCTGGGCTG 1180
DB 363 GGCAGGTGGAAATGACGACGCTCTGCCAGAGGCCCTACACTGGGTCTGCGAGGCTGGGCTG 304
QY 1181 GGTTCAGACAGCCAGGAGAGTCACTCAGCTGCCTTTGGTGGGACCAACCCGGCCACAGAAA 1240
DB 303 GGTTCAGACAGCCAGGAGAGTCACTCAGCTGCCTTTGGTGGGACCAACCCGGCCACAGAAA 244
QY 1241 TGGCGGTGGAGGAGGAGTCTTCTCAGCAGCTCTCTCGCAGACCGCTCTGGGAGAGAAAT 1300
DB 243 TGGCGGTGGAGGAGGAGTCTTCTCAGCAGCTCTCTCGCAGACCGCTCTGGGAGAGAAAT 184
QY 1301 AAGCAGTGGAGATTGGAAGCACTGCTAACATTTTGAATTTTCTCTTTAATTTTAAA 1360
DB 183 AAGCAGTGGAGATTGGAAGCACTGCTAACATTTTGAATTTTCTCTTTAATTTTAAA 124
QY 1361 AAGATGATATAGTGTCTTAAAGCTTTTATTTTTCCTCAACTTTTGAAAGTCAACTTCAT 1420
DB 123 AAGATGATATAGTGTCTTAAAGCTTTTATTTTTCCTCAACTTTTGAAAGTCAACTTCAT 64
QY 1421 GAAGGTATATATTTTACATAATAAATAATGCACCTCATTT 1458
DB 63 GAAGGTATATATTTTACATAATAAATAATGCACCTCATTT 26

BI905917 634 bp mRNA linear EST 16-OCT-2001
603063080F1 NIH_MGC_118 Homo sapiens cDNA clone IMAGE:5212290 5',
mRNA sequence.
BI905917
VERSION BI905917.1 GI:16168336
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 634)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-@email.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM1532 row: p column: 19
High quality sequence stop: 632.
Location/Qualifiers
1..634
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5212290"
/tissue_type="leukocyte"
/lab_host="DH10B"
/clone_lib="NIH_MGC_118"
/notes="Vector: pCMV-SPORT6; Site 1: NotI; Site 2: EcorV
(destroyed); RNA source leukocytes from anonymous pool of
non-activated adult donors. Library is oligo-dT primed
and directionally cloned (EcorV site is destroyed upon
cloning). Average insert size 1.7 Kb, insert size range

1.2-3.3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 027. Note: this is a NIH_MGC Library."

ORIGIN		Query Match	42.6%;	Score 621.4;	DB 4;	Length 634;	
		Best Local Similarity	99.7%;	Pred. No. 7.2e-151;			
		Mismatches	633;	Conservative	0;	Mismatches	1; Indels 1; Gaps 1;
QY	10	ATGGGATGCTCCAGATGATAGGCTCTCGGATTTCCAGACCCAGACAGGAGCTCCA	69				
DB	1	ATGGGATGCTCCAGATGATAGGCTCTCGGATTTCCAGACCCAGACAGGAGCTCCA	60				
QY	70	GTCACTCTACCCAGCTCTCCAGGACACAGGCTCCCAACTCTGAGTCAGCTCCCACT	129				
DB	61	GTCACTCTACCCAGCTCTCCAGGACACAGGCTCCCAACTCTGAGTCAGCTCCCACT	120				
QY	130	CTGGTCTTTGACGACAAACAGCTGGGAATCACACCTCCAGACCTCCCAAGCTCCAC	189				
DB	121	CTGGTCTTTGACGACAAACAGCTGGGAATCACACCTCCAGACCTCCCAAGCTCCAC	180				
QY	190	CCAGACTGGGGCGCGGCTGCTCCATTTGAGCTGACAACTCTGAGCGGCTGG	249				
DB	181	CCAGACTGGGGCGCGGCTGCTCCATTTGAGCTGACAACTCTGAGCGGCTGG	240				
QY	250	CCCAAGCATGACAAAGGACGTATGAAACTTTCCAGTACTTTGGGAATAAGGTGAAAGTCCA	309				
DB	241	CCCAAGCATGACAAAGGACGTATGAAACTTTCCAGTACTTTGGGAATAAGGTGAAAGTCCA	300				
QY	310	GGGGTTTAAATATGGGCCATTTCTCTCCAGTCCCTCTCGAGCGCTCTCCGCTCTGGGCC	369				
DB	301	-GGGTTTAAATATGGGCCATTTCTCTCCAGTCCCTCTCGAGCGCTCTCCGCTCTGGGCC	359				
QY	370	CTGCCATCTCTGCTGCTCTCGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	429				
DB	360	CTGCCATCTCTGCTGCTCTCGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	419				
QY	430	TGGATTCCAAATTTCCAAATTTCCAGAGGACCTGGTGACCTCCAGAACAGATTTTAGCAA	489				
DB	420	TGGATTCCAAATTTCCAAATTTCCAGAGGACCTGGTGACCTCCAGAACAGATTTTAGCAA	479				
QY	490	CTTCACTCCAAACACTGTGGCGAGATCCAGGCACCTGACTTCCAGGGCAGCAGCTTGA	549				
DB	480	CTTCACTCCAAACACTGTGGCGAGATCCAGGCACCTGACTTCCAGGGCAGCAGCTTGA	539				
QY	550	AGAAACGATAGCATCTCTGAAGCTCAGGTGGAGGTTTCAAGCAGGAACGGCAGG	609				
DB	540	AGAAACGATAGCATCTCTGAAGCTCAGGTGGAGGTTTCAAGCAGGAACGGCAGG	599				
QY	610	GGTATCTGAGCTCCAGGAACACACTACGACAGG	644				
DB	600	GGTATCTGAGCTCCAGGAACACACTACGACAGG	634				

RESULT 15	BI767532	816 bp	mRNA	linear	EST 25-SEP-2001	
LOCUS	603061425F1	NIH_MGC_122	Homo sapiens	cdna clone	IMAGE:5210406 5',	
DEFINITION	mRNA sequence.					
ACCESSION	BI767532					
VERSION	BI767532.1	GI:15759110				
KEYWORDS	EST.					
SOURCE	Homo sapiens (human)					
ORGANISM	Homo sapiens					
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
REFERENCE	1	(bases 1 to 816)				
AUTHORS	NIH-MGC	http://mgi.nci.nih.gov/ .				
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)					
JOURNAL	Unpublished (1999)					
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgabbs-remail.nih.gov					

Tissue Procurement: Life Technologies, Inc.
cdna Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM11528 row: b column: 07
High quality sequence start: 2
High quality sequence stop: 812.
Location/Qualifiers
1. 816
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5210406"
/lab_host="DH10B"
/clone_lib="NIH_MGC_122"
/note="Organ: pooled lung and spleen; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: EcoRV (destroyed); RNA source anonymous pool of 24 week female lung, 16 week female spleen, and 20-22 week male spleens. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.4 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 026. Note: this is a NIH_MGC Library."

FEATURES
source

ORIGIN

Query Match 42.3%; Score 616.6; DB 4; Length 816;
Best Local Similarity 88.7%; Pred. No. 1.4e-149;
Matches 773; Conservative 0; Mismatches 4; Indels 94; Gaps 6;

QY	1	GTGAGGAGATGGGATGTCCTCCAGATGATAGGCTCTCGGATTTCCAGACCCAGACGAC	60
DB	24	GATGAGGAGATGGGATGTCCTCCAGATGATAGGCTCTCGGATTTCCAGACCCAGACGAC	83
QY	61	AGGACTCCAGTCACCTACCCAGCTCTCCAGGACACAGCGCTCCCACTCTGA-GTGA	119
DB	84	AGGACTCCAGTCACCTACCCAGCTCTCCAGGACACAGCGCTCCCACTCTGAGGTGA	143
QY	120	CGTCCACCTCTGCTGCTTTGAGCACAACCAACGTTGGGAATCACACCTCCAGACCTCCC	179
DB	144	CGTCCACCTCTGCTGCTTTGAGCACAACCAACGTTGGGAATCACACCTCCAGACCTCCC	203
QY	180	ACAGTCCACCCAGACAGCTGGGGCGGCGCTGCTCCATTTTCAGCTGTGACACCTCAGA	239
DB	204	ACAGTCCACCCAGACAGCTGGGGCGGCGCTGCTCCATTTTCAGCTGTGACACCTCAGA	263
QY	240	GCCGTTGGCCCAAGCATGACAGGACGTATGAAAACCTTCAGTACTTGGAGAATAAGG	299
DB	264	GCCGTTGGCCCAAGCATGACAGGACGTATGAAAACCTTCAGTACTTGGAGAATAAGG	323
QY	300	TGAAAGTCCAGGGGTTTAAAAATGGGCCACTTCTCTCCAGTCCCTCTCCAGCGCTCTCC	359
DB	324	TGAAAGTCCAGGGGTTTAAAAATGGGCCACTTCTCTCCAGTCCCTCTCCAGCGCTCTCT	383
QY	360	GCTCTGGGCGCTGCCATCTCTGCTGCTGCTGGGCTCGGCTGCTGCTGCTGCTGCTGCTCA	419
DB	384	GCTCTGGGCGCTGCCATCTCTGCTGCTGCTGGGCTCGGCTGCTGCTGCTGCTGCTGCTCA	443
QY	420	TCTGTGTGTTGGATTCCAAATTTCCAAATTTTCAGAGGACCTGTGTGACCTTGAGAACAG	479
DB	444	TCTGTGTGTTGGATTCCAAATTTCCAAATTTTCAGAGGACCTGTGTGACCTTGAGAACAG	503
QY	480	ATTTTAGCAACTTCACTCAAA-CAGTGTGGCGGAGATCCAGGCACTGACTTCCAGGGC	538
DB	504	ATTTTAGCAACTTCACTCAAAAGCACTGTGGCGGAGATCCAGGCACTGACTTCCAGGGC	563
QY	539	AGCAGCTTGGAGAAGAACGATAGCATCTCTGAAGCTGAGGTGGAGGGTTTCAGCAGGNA	598
DB	564	AGCAGCTTGGAGAAGAACGATAGCATCTCTGAAGCTGAGGTGGAGGGTTTCAGCAGGNA	623

```
Qy 599 CGGCGGCGAGGGGTATCTGAGCTCCAGGAACACTAGGAGAGGCACACTAGGCCAC 658
Db      |||||
624 CGGAGG----- 630
Qy 659 TGTCCCCACTGCCCATCTGTGTGTGCCAGTTCAATTCTGAATGCTCTCGGAGTCCAG 718
Db      |||||
631 -----CAGTTCAATCTGAATGCTCTCGGAGTCCAG 662
Qy 719 CAGCTGTGTC-AAGACCTGAAGAAACTGACCTGCCAGGTGGCTACTCTCAACAACAAT-- 775
Db      |||||
663 CAGCTGTGCAAGAACCTGAAGAAACTGACCTGCCAGGTGGCTACTCTCAACAACAATGG 722
Qy 776 -----GCCTCCACTGAAGGGACCTGCTGCCCGTCAACTGGGTGGAGCACCAGACAG 828
Db      |||||
723 TGAGGAAGCCTCCACTGAAGGGACCTGCTGCCCTGTCACTGGGTGGAGCACCAGACAG 782
Qy 829 CTGCTACTGGTTC-TCTCACTCTGGGATGTC 858
Db      |||||
783 CTGCTACTGGTTCCTCTCACTCTGGGATGTC 813
```

Search completed: May 27, 2005, 23:52:45
Job time : 5121 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 27, 2005, 19:12:00 ; Search time 6433 Seconds
(without alignments)
10982.078 Million cell updates/sec

Title: US-10-829-107-3
Perfect score: 1458
Sequence: 1 gttgagagatggatgtcc.....taataaaatgcactcattt 1458

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*
1: gb_ba.*
2: gb_hcg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sta.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1458	100.0	1458	6	BD135981 Isolated
2	1458	100.0	1458	6	AR561599 Sequence
3	1446.6	99.2	1769	9	BC039011 Homo sapi
4	1218	83.5	1344	9	DS0532 Homo sapien
5	1108	76.0	1370	6	BD135983 Isolated
6	1108	76.0	1370	6	AR561601 Sequence
7	857.4	58.8	1858	9	BC027858 Homo sapi
8	606.6	41.6	661	6	AX067341 Sequence
9	438	30.0	1358	10	RAT03GSLC
10	420.8	28.9	172095	2	AC026834 Homo sapi
11	420.8	28.9	176813	2	AC107895 Homo sapi
12	420.8	28.9	198821	9	AC120057 Homo sapi
13	420.8	28.9	220581	2	AC015918 Homo sapi
14	404.6	27.8	1414	10	S36676 galactose a
15	403	27.6	1353	10	BC014811 Mus muscu
16	399.8	27.4	1520	10	AY103461 Mus muscu
17	382	26.2	1196	9	BC032130 Homo sapi
18	382	26.2	1277	6	E12702 cDNA encodi
19	382	26.2	1277	6	AR270623 Sequence

20	382	26.2	1277	6	AX409583 Sequence
21	382	26.2	1277	9	HUMASGPR1
22	382	26.2	1657	9	AB070933 Homo sapi
23	374.6	25.7	873	9	CR542052 Homo sapi
24	370	25.4	1223	6	CQ725393 Sequence
25	344.2	23.6	853	11	BV208440 CLECSF14
26	344.2	23.6	1232	10	BC022106 Mus muscu
27	335.6	23.0	1328	10	MUSMHL1
28	334	22.9	991	10	MMU09362
29	329	22.6	1309	10	MMU08372
30	321.4	22.0	1300	6	E12703
31	321.4	22.0	1300	6	AX411087
32	321.4	22.0	1300	9	HSLH2MR
33	321.4	22.0	1344	9	BC017251
34	320.2	22.0	971	10	RATASGP2
35	310.4	21.3	361	6	AX897915
36	310.4	21.3	361	6	BD033448 Sequence
37	307.8	21.1	921	9	AF529374 Homo sapi
38	290.6	19.9	879	9	HSU97197 Homo sapien
39	283.8	19.5	1308	6	CQ725392 Sequence
40	282.8	19.4	1309	6	AX409588 Sequence
41	282.8	19.4	1309	6	AX774853 Sequence
42	282.8	19.4	1309	9	HUMASGPR2
43	273.4	18.8	1382	9	HSLH2BMR
44	267	18.3	428	10	AY643476 Marmota m
45	263	18.0	1470	10	AF230645 Rattus no

ALIGNMENTS

RESULT 1

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BD135981 1458 bp DNA linear PAT 18-SEP-2002
Isolated dentritic cell membrane protein genes.
BD135981
BD135981.1 GI:23230926
JP 2002509438-A/2.
unidentified
unclassified.
1 (bases 1 to 1458)
Valladeau, J., Ravel, O., Bates, E.E.M., Ford, J., Saeland, S. and Lebecque, S.J.E.
Isolated dentritic cell membrane protein genes
Patent: JP 2002509438-A 2 26-MAR-2002;
SCHERING CORP
OS Unidentified
PN JP 2002509438-A/2
PD 26-MAR-2002
PF 08-JUL-1998 JP 1999508710
PR 09-JUL-1997 US 60/053080
PI JENNY VALLADEAU, ODILE RAVEL, ELIZABETH ESTHER MARY BATES, JOHN FORD.
PI SEM SÆLAND, SERGE J E LEBECQUE
PC C07K14/705, C12N15/12
CC Strandedness: Single;
CC Topology: Linear;
CC /note= 'short form lacks nucleotides 608-673' CC /note= 'ASGPRm (table 2) has sequence insert encoding GEE CC between CC nucleotides 775-776'
CC /note= 'nucleotide 1064 of DCMF2s may be A, which would encode asn rather CC than asp at the residue numbered 270'
FH Key Location/Qualifiers
FT CDS 257..1204
FT misc feature 608
FT misc feature 775
FT misc feature 1064.
FT Location/Qualifiers
source 1..1458

```
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"

ORIGIN
Query Match      100.0%; Score 1458; DB 6; Length 1458;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1458; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTTGGAGAGATGGGATGTCCTCCAGATGATAGGGCTCTCTGGGATTTTCAGACCCCAAGACCAGC 60
DB |||||||
QY 1 GTTGGAGAGATGGGATGTCCTCCAGATGATAGGGCTCTCTGGGATTTTCAGACCCCAAGACCAGC 60
DB |||||||
QY 61 AGGACTCCAGTCACTCTACCCAGCTCTCCAGGACACAGCGCTCCCAACTCTGAGTGAC 120
DB |||||||
QY 61 AGGACTCCAGTCACTCTACCCAGCTCTCCAGGACACAGCGCTCCCAACTCTGAGTGAC 120
DB |||||||
QY 121 GTCCCACTCTGGTCTCTGAGACCAACCAAGTGGGAATCAACCTCCAGACCTCCCA 180
DB |||||||
QY 121 GTCCCACTCTGGTCTCTGAGACCAACCAAGTGGGAATCAACCTCCAGACCTCCCA 180
DB |||||||
QY 181 CAGCTCCACCCAGACTGGGCGCGCCCTGCTCCATTTCAAGCTGTGACAACTCAGAG 240
DB |||||||
QY 181 CAGCTCCACCCAGACTGGGCGCGCCCTGCTCCATTTCAAGCTGTGACAACTCAGAG 240
DB |||||||
QY 241 CCGTGTGGCCCAAGCATGACAAAGGACGTATGAAAACCTTCCAGTACTTTGGAGAAATAAGGT 300
DB |||||||
QY 241 CCGTGTGGCCCAAGCATGACAAAGGACGTATGAAAACCTTCCAGTACTTTGGAGAAATAAGGT 300
DB |||||||
QY 301 GAAAGTCCAGGGGTTTAAAAATGGGCACTTCTCTCAGTCCCTCTCGACGGCTCTCG 360
DB |||||||
QY 301 GAAAGTCCAGGGGTTTAAAAATGGGCACTTCTCTCAGTCCCTCTCGACGGCTCTCG 360
DB |||||||
QY 361 CTCCTGGGCGCTGCATCTCTGCTGTCTGGGCGCTGGGCGCTGGGCGCTGGGCGCTGGGCGCT 420
DB |||||||
QY 361 CTCCTGGGCGCTGCATCTCTGCTGTCTGGGCGCTGGGCGCTGGGCGCTGGGCGCTGGGCGCT 420
DB |||||||
QY 421 CTGTGTGGTGGATTCAAAATTTCCAAAATTTTCAGAGGGACCTGGTGACCTTCAGAACAGA 480
DB |||||||
QY 421 CTGTGTGGTGGATTCAAAATTTCCAAAATTTTCAGAGGGACCTGGTGACCTTCAGAACAGA 480
DB |||||||
QY 481 TTTTAGCAATCTCACTCAAACTGTGGGGAGATCCAGGACCTGACCTTCACAGGGGAG 540
DB |||||||
QY 481 TTTTAGCAATCTCACTCAAACTGTGGGGAGATCCAGGACCTGACCTTCACAGGGGAG 540
DB |||||||
QY 541 CAGCTTGGAGAAACCATAGCATCTCTGAAAGCTGAGGTGGAGGTTTCAACGACGAAGC 600
DB |||||||
QY 541 CAGCTTGGAGAAACCATAGCATCTCTGAAAGCTGAGGTGGAGGTTTCAACGACGAAGC 600
DB |||||||
QY 601 GCAGGCAGGGGTATCTGAGCTCCAGGAACACACTACGCAAGAGGCACACTTAGGCCACTG 660
DB |||||||
QY 601 GCAGGCAGGGGTATCTGAGCTCCAGGAACACACTACGCAAGAGGCACACTTAGGCCACTG 660
DB |||||||
QY 661 TCCCACTGCCCACCTGTGTGTGTCCTGAAATGCTCTGAAATGCTCTCGAGTCCAGCA 720
DB |||||||
QY 661 TCCCACTGCCCACCTGTGTGTGTCCTGAAATGCTCTGAAATGCTCTCGAGTCCAGCA 720
DB |||||||
QY 721 GCTGTGTGCAAGACCTCAAGAACTGACCTGCCAGGTGGCTACTCTCAACAACTATGCTCT 780
DB |||||||
QY 721 GCTGTGTGCAAGACCTCAAGAACTGACCTGCCAGGTGGCTACTCTCTCAACAACTATGCTCT 780
DB |||||||
QY 781 CACTGAAGGGACCTGTGCTGCCCCCTCAACTGGGTGGAGCAACCAAGCAGCTGTACTGTGTT 840
DB |||||||
QY 781 CACTGAAGGGACCTGTGCTGCCCCCTCAACTGGGTGGAGCAACCAAGCAGCTGTACTGTGTT 840
DB |||||||
QY 841 CTCTCACTCTGGGATGCTCTGGGCGGAGGTGAGAAATGCTGCGAGCTGAAGAAAGCCCA 900
DB |||||||
QY 841 CTCTCACTCTGGGATGCTCTGGGCGGAGGTGAGAAATGCTGCGAGCTGAAGAAAGCCCA 900
DB |||||||
QY 901 CTTGTGGTCACTCACTCCAGGGAGGAGCAGAAATTTTGTCCAGAAATATCTAGGCTCCGC 960
DB |||||||
QY 901 CTTGTGGTCACTCACTCCAGGGAGGAGCAGAAATTTTGTCCAGAAATATCTAGGCTCCGC 960
DB |||||||

RESULT 2
AR561599
LOCUS AR561599 1458 bp DNA linear PAT 08-OCT-2004
DEFINITION Sequence 3 from patent US 6756478.
ACCESSION AR561599
VERSION AR561599.1 GI:53974518
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1458)
AUTHORS Valladeau, J., Ravel, O., Bates, E.B.M., Ford, J., Saeland, S. and Lebecque, S.J.E.
TITLE Isolated mammalian membrane protein genes; related reagents
JOURNAL Patent: US 6756478-A 3 29-JUN-2004;
FEATURES
source location/Qualifiers
1..1458
/mol_type="genomic DNA"

ORIGIN
Query Match      100.0%; Score 1458; DB 6; Length 1458;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1458; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTTGGAGAGATGGGATGTCCTCCAGATGATAGGGCTCTCTGGGATTTTCAGACCCCAAGACCAGC 60
DB |||||||
QY 1 GTTGGAGAGATGGGATGTCCTCCAGATGATAGGGCTCTCTGGGATTTTCAGACCCCAAGACCAGC 60
DB |||||||
QY 61 AGGACTCCAGTCACTCTACCCAGCTCTCCAGGACACAGCGCTCCCAACTCTGAGTGAC 120
DB |||||||
QY 61 AGGACTCCAGTCACTCTACCCAGCTCTCCAGGACACAGCGCTCCCAACTCTGAGTGAC 120
DB |||||||
QY 121 GTCCCACTCTGGTCTCTGAGACCAACCAAGTGGGAATCAACCTCCAGACCTCCCA 180
DB |||||||
QY 121 GTCCCACTCTGGTCTCTGAGACCAACCAAGTGGGAATCAACCTCCAGACCTCCCA 180
DB |||||||
QY 181 CAGCTCCACCCAGACTGGGCGCGCCCTGCTCCATTTCAAGCTGTGACAACTCAGAG 240
DB |||||||
QY 181 CAGCTCCACCCAGACTGGGCGCGCCCTGCTCCATTTCAAGCTGTGACAACTCAGAG 240
DB |||||||
QY 241 CCGTGTGGCCCAAGCATGACAAAGGACGTATGAAAACCTTCCAGTACTTTGGAGAAATAAGGT 300
DB |||||||
QY 241 CCGTGTGGCCCAAGCATGACAAAGGACGTATGAAAACCTTCCAGTACTTTGGAGAAATAAGGT 300
DB |||||||
QY 301 GAAAGTCCAGGGGTTTAAAAATGGGCACTTCTCTCAGTCCCTCTCGACGGCTCTCG 360
DB |||||||
QY 301 GAAAGTCCAGGGGTTTAAAAATGGGCACTTCTCTCAGTCCCTCTCGACGGCTCTCG 360
DB |||||||
QY 361 CTCCTGGGCGCTGCATCTCTGCTGTCTGGGCGCTGGGCGCTGGGCGCTGGGCGCTGGGCGCT 420
DB |||||||
QY 361 CTCCTGGGCGCTGCATCTCTGCTGTCTGGGCGCTGGGCGCTGGGCGCTGGGCGCTGGGCGCT 420
DB |||||||
QY 421 CTGTGTGGTGGATTCAAAATTTCCAAAATTTTCAGAGGGACCTGGTGACCTTCAGAACAGA 480
DB |||||||
QY 421 CTGTGTGGTGGATTCAAAATTTCCAAAATTTTCAGAGGGACCTGGTGACCTTCAGAACAGA 480
DB |||||||
QY 481 TTTTAGCAATCTCACTCAAACTGTGGGGAGATCCAGGACCTGACCTTCACAGGGGAG 540
DB |||||||
QY 481 TTTTAGCAATCTCACTCAAACTGTGGGGAGATCCAGGACCTGACCTTCACAGGGGAG 540
DB |||||||
QY 541 CAGCTTGGAGAAACCATAGCATCTCTGAAAGCTGAGGTGGAGGTTTCAACGACGAAGC 600
DB |||||||
QY 541 CAGCTTGGAGAAACCATAGCATCTCTGAAAGCTGAGGTGGAGGTTTCAACGACGAAGC 600
DB |||||||
QY 601 GCAGGCAGGGGTATCTGAGCTCCAGGAACACACTACGCAAGAGGCACACTTAGGCCACTG 660
DB |||||||
QY 601 GCAGGCAGGGGTATCTGAGCTCCAGGAACACACTACGCAAGAGGCACACTTAGGCCACTG 660
DB |||||||
QY 661 TCCCACTGCCCACCTGTGTGTGTCCTGAAATGCTCTGAAATGCTCTCGAGTCCAGCA 720
DB |||||||
QY 661 TCCCACTGCCCACCTGTGTGTGTCCTGAAATGCTCTGAAATGCTCTCGAGTCCAGCA 720
DB |||||||
QY 721 GCTGTGTGCAAGACCTCAAGAACTGACCTGCCAGGTGGCTACTCTCAACAACTATGCTCT 780
DB |||||||
QY 721 GCTGTGTGCAAGACCTCAAGAACTGACCTGCCAGGTGGCTACTCTCTCAACAACTATGCTCT 780
DB |||||||
QY 781 CACTGAAGGGACCTGTGCTGCCCCCTCAACTGGGTGGAGCAACCAAGCAGCTGTACTGTGTT 840
DB |||||||
QY 781 CACTGAAGGGACCTGTGCTGCCCCCTCAACTGGGTGGAGCAACCAAGCAGCTGTACTGTGTT 840
DB |||||||
QY 841 CTCTCACTCTGGGATGCTCTGGGCGGAGGTGAGAAATGCTGCGAGCTGAAGAAAGCCCA 900
DB |||||||
QY 841 CTCTCACTCTGGGATGCTCTGGGCGGAGGTGAGAAATGCTGCGAGCTGAAGAAAGCCCA 900
DB |||||||
QY 901 CTTGTGGTCACTCACTCCAGGGAGGAGCAGAAATTTTGTCCAGAAATATCTAGGCTCCGC 960
DB |||||||
QY 901 CTTGTGGTCACTCACTCCAGGGAGGAGCAGAAATTTTGTCCAGAAATATCTAGGCTCCGC 960
DB |||||||
```


QY 181 CAGTCCACCCAGACCTGGGGCGGGCCCTGCTCCATTTTCAGTGTGACAACTCAGAG 240
Db 181 CAGTCCACCCAGACCTGGGGCGGGCCCTGCTCCATTTTCAGTGTGACAACTCAGAG 240
QY 241 CCGTGTGGCCCAAGCATGACAGGACGTATGAAAACCTTCAGTACTTGGAGAAATAGGT 300
Db 241 CCGTGTGGCCCAAGCATGACAGGACGTATGAAAACCTTCAGTACTTGGAGAAATAGGT 300
QY 301 GAAAGTCAGGGGTTTAAAAATGGCCACTTCTCTCTCCAGTCCCTCTCGAGGCTCTCCG 360
Db 301 GAAAGTCAGGGGTTTAAAAATGGCCACTTCTCTCTCCAGTCCCTCTCGAGGCTCTCCG 360
QY 361 CTCTGGGCGCTGCATCTCTGCTGCTCCCTGGGCGCTGCTGCTGCTGCTGCTGCTCAT 420
Db 361 CTCTGGGCGCTGCATCTCTGCTGCTCCCTGGGCGCTGCTGCTGCTGCTGCTGCTCAT 420
QY 421 CTCTGTGTTGGATTCCAAAATTTCCAAATTTTCAGAGGGACCTGTGTGACCTCTGAGAACAGA 480
Db 421 CTCTGTGTTGGATTCCAAAATTTCCAAATTTTCAGAGGGACCTGTGTGACCTCTGAGAACAGA 480
QY 481 TTTTAGCACTTCACTCAAAACACTGTGGCGGAGATCCAGGCACTGACTTTCCAGGGCAG 540
Db 481 TTTTAGCACTTCACTCAAAACACTGTGGCGGAGATCCAGGCACTGACTTTCCAGGGCAG 540
QY 541 CAGTTGGAAGAAACGATGATCTCTGAAAGCTGAGGTGGAGGGTTTCAAGCAGGAACG 600
Db 541 CAGTTGGAAGAAACGATGATCTCTGAAAGCTGAGGTGGAGGGTTTCAAGCAGGAACG 600
QY 601 GCAGGCGAGGGTATCTGAGCTCCAGGAACACTAGCAGCAAGGACACACCTTAGGCCACTG 660
Db 601 GCAGGCGAGGGTATCTGAGCTCCAGGAACACTAGCAGCAAGGACACACCTTAGGCCACTG 660
QY 661 TCCCACTGCCCCTCTGTGTGTGTCAGGTTTCAATTTCTGAAATGCTCTCTGCGAGTCCAGCA 720
Db 661 TCCCACTGCCCCTCTGTGTGTGTCAGGTTTCAATTTCTGAAATGCTCTCTGCGAGTCCAGCA 720
QY 721 GCTGTGTCAGACCTGAGAAACCTGACCTGCCAGGTGGCTACTCTCAACAACTATGCTC 780
Db 721 GCTGTGTCAGACCTGAGAAACCTGACCTGCCAGGTGGCTACTCTCAACAACTATGCTC 780
QY 781 CACTGAAGGACCTGCTGCCCCCTCACTGCGTGGAGCACCAGACAGCTGCTACTGGTT 840
Db 781 CACTGAAGGACCTGCTGCCCCCTCACTGCGTGGAGCACCAGACAGCTGCTACTGGTT 840
QY 841 CTCTCACTCTGGGATGCTCTGGGCGGAGGCTGAGAGTACTGCCAGCTGAAGAACGCCCA 900
Db 841 CTCTCACTCTGGGATGCTCTGGGCGGAGGCTGAGAGTACTGCCAGCTGAAGAACGCCCA 900
QY 901 CTTGTGTTGTCATCAACTCCAGGGAGGAGCAGAAATTTTGTCCAGAAATATCTAGGCTCCG 960
Db 901 CTTGTGTTGTCATCAACTCCAGGGAGGAGCAGAAATTTTGTCCAGAAATATCTAGGCTCCG 960
QY 961 ATACACTGATGGGCTCAGTGCCTGAGGAGCCTGAGGAGCTGGTGGATGGAACAGA 1020
Db 961 ATACACTGATGGGCTCAGTGCCTGAGGAGCCTGAGGAGCTGGTGGATGGAACAGA 1020
QY 1021 CTATGCCAGCGGCTTCAGAACTGGAAGCCAGGCGCCAGACGACTGGCAGGGGACCGG 1080
Db 1021 CTATGCCAGCGGCTTCAGAACTGGAAGCCAGGCGCCAGACGACTGGCAGGGGACCGG 1080
QY 1081 GCTGGGTGGAGGGAGGAGCTGTGCTCACTTCCATCCAGACGGCAGGTGGAATGACGAGT 1140
Db 1081 GCTGGGTGGAGGGAGGAGCTGTGCTCACTTCCATCCAGACGGCAGGTGGAATGACGAGT 1140
QY 1141 CTGCCAGAGGCGCTTACCACTGGGTCTCGAGGCTGGGCTGGTCCAGACCCAGCAGGAG 1200
Db 1141 CTGCCAGAGGCGCTTACCACTGGGTCTCGAGGCTGGGCTGGTCCAGACCCAGCAGGAG 1200
QY 1201 TCACTGAGCTGCTTTGTTGGGACACCCGSCCAAGAAATGGCGGTGGGAGGAGGACTC 1260
Db 1201 TCACTGAGCTGCTTTGTTGGGACACCCGSCCAAGAAATGGCGGTGGGAGGAGGACTC 1260

QY 1261 TTCTCAGACCTCTCGAAGACCGCTCTGGGAGAGAAATAAGCACTGGGAGATTGGAAG 1320
Db 1261 TTCTCAGACCTCTCGAAGACCGCTCTGGGAGAGAAATAAGCACTGGGAGATTGGAAG 1320
QY 1321 CACTGCTAACATTTTGAATTTTTTTCTCTTTAAATTTTAAAGATGGTATAGTGTCTTA 1380
Db 1321 CACTGCTAACATTTTGAATTTTTTTCTCTTTAAATTTTAAAGATGGTATAGTGTCTTA 1380
QY 1381 AGCTTTTATTTTTTTTCCAACTTTTGAAGTCAACTTCATGAAAGGTATAATTTTACATA 1440
Db 1381 AGCTTTTATTTTTTTTCCAACTTTTGAAGTCAACTTCATGAAAGGTATAATTTTACATA 1440
QY 1441 ATAAAAATGCACATTT 1458
Db 1441 ATAAAAATGCACATTT 1458

BC039011 1769 bp mRNA linear PRI 30-JUN-2004
Homo sapiens C-type (calcium dependent, carbohydrate-recognition domain) lectin, superfamily member 14 (macrophage-derived), transcript variant 1, mRNA (cDNA clone MGC:47699 IMAGE:5751735), complete cds.

BC039011 GI:24660336
MGC.
Homo sapiens (human)
Homo sapiens
ORGANISM

REFERENCE
AUTHORS
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1769)
Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Spletton,T.E., Soares,M.B., Bonaldo,M.P., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Kettman,M., Madan,A., Rodriguez,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakeley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smalios,D.E., Schnerch,A., Schein,J.E., Jones,S.U. and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

12477932
2 (bases 1 to 1769)
Strausberg,R.
Direct Submission
Submitted (01-NOV-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

TITLE
JOURNAL
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL

REMARK
COMMENT
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: <http://www.nisc.nih.gov/>
Contact: nisc_mgc@hgrl.nih.gov
Akhtar,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,

Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R., Maduro, O.L., Masiello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W., Tsourgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILNLT at: <http://image.llnl.gov>
 Series: IRAX Plate: 82 Row: j Column: 7
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 5453683.

FEATURES

source

Location/Qualifiers
 1. .1769
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="MGC:47699 IMAGE:5751735"
 /tissue_type="Brain, Lung, Testis, adult, pooled whole"
 /clone_lib="NIH MGC_115"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"

gene

1. .1769
 /gene="CLECSF14"
 /note="synonyms: HML2, HML, CLECSF13"
 /db_xref="LocustID:10462"
 /db_xref="MIM:605999"

CDS

252..1202
 /gene="CLECSF14"
 /codon_start=1
 /product="C-type lectin, superfamily member 13, isoform 1"
 /protein_id="AAH39011.1"
 /db_xref="GI:24660337"
 /db_xref="LocustID:10462"
 /db_xref="MIM:605999"
 /translation="MTRTYENFYLENKVKVQFKNGPLQLSLRLQCSGPHLLLS
 LGILLILVVICVGFQNSKFORDLVLTDFSNFTSVTAIEIOALTSGSLEBETIA
 SLKAEVGFGRORAGVSELOBHTTQKAHLGHCPHCPVSVVPHSEMILLRVOQLVODL
 KKLTCQVATLNNASTEGTCCPVNVEHODSCWYFESHSGMSWAEKVCQLKNAHLV
 INSRERQNFQKILGSAIYTWGLSDPEGAWKWDGDIYATGFQNKWPGQPDWQGHGL
 GGGEDCAHPHPDGRWNDDVCQRPYHVVCEAGLGQTSQESH"

ORIGIN

Query Match 99.2%; Score 1446.6; DB 9; Length 1769;
 Best Local Similarity 99.7%; Pred. No. 0;
 Matches 1449; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy	6	GGAGATGGGATGTCCAGATGATAGGGCTCCTGGGATTTCCAGACCCAGACGAGGAC	65
Db	1	GGAGATGGGATGTCCAGATGATAGGGCTCCTGGGATTTCCAGACCCAGACGAGGAC	60
Qy	66	TCAGTCACTCTACCCAGCTCTCAGAGACAGCGCTCCCAACTCTGAGTGAGTCCCT	125
Db	61	TCAGTCACTCTACCCAGCTCTCAGAGACAGCGCTCCCAACTCTGAGTGAGTCCCT	120
Qy	126	ACCTCTGGTCTTGACAGCAACCAACGTTGGGAATCACACCTCCAGACCTCCACAGCT	185
Db	121	ACCTCTGGTCTTGACAGCAACCAACGTTGGGAATCACACCTCCAGACCTCCACAGCT	180
Qy	186	CCACCCAGATGGGGCCGGCCCTCCCTCCATTTTCAGCTGTGACAACTCCAGACCGTG	245
Db	181	CCACCCAGATGGGGCCGGCCCTCCCTCCATTTTCAGCTGTGACAACTCCAGACCGTG	240
Qy	246	TTGGCCCAAGCATGACAAGAGAGTATGAAAACTTCCAGTACTTGGAGAAATAGGTGAAG	305
Db	241	TTGGCCCAAGCATGACAAGAGAGTATGAAAACTTCCAGTACTTGGAGAAATAGGTGAAG	300
Qy	306	TCAGGGGTTTAAAAATGGCCACTTCTCTCCAGTCCCTCTCCAGCGCTCTCCGCTCTG	365
Db	301	TCAGGGGTTTAAAAATGGCCACTTCTCTCTCCAGTCCCTCTCCAGCGCTCTCTGCTCTG	360
Qy	366	GGCCCTGCCATCTCTCTGCTGTCCCTGGGCTCGGCCCTGCTGCTGCTCATCATCTGTG	425

Db	361	GGCCCTGCCATCTCTGCTGTCTGCTGGCCCTCGCCCTCTCTGCTGCTCATCATCTGTG	420
Qy	426	TGTTTGGATTCAAAATTCAAAATTCAGAGGGACCTGGTACACCTGAGAACAGATTTTA	485
Db	421	TGTTTGGATTCAAAATTCAAAATTCAGAGGGACCTGGTACACCTGAGAACAGATTTTA	480
Qy	486	GCAACTTCACCTCAAAACACTGTGGGGAGATCCAGGCACTGACTTCCAGGGCAGCAGCT	545
Db	481	GCAACTTCACCTCAAAACACTGTGGGGAGATCCAGGCACTGACTTCCAGGGCAGCAGCT	540
Qy	546	TGGAGAAAACGATAGCATCTCTGAAAAGCTGAGGTGGAGGGTTTCAAGCAGAAACGGCAGG	605
Db	541	TGGAGAAAACGATAGCATCTCTGAAAAGCTGAGGTGGAGGGTTTCAAGCAGAAACGGCAGG	600
Qy	606	CAGGGTATCTGAGCTCCAGAACACACTAGCCAGAGGACACCTAGGCCACTGTCCCC	665
Db	601	CAGGGTATCTGAGCTCCAGAACACACTAGCCAGAGGACACCTAGGCCACTGTCCCC	660
Qy	666	ACTGCCCATCTGT	725
Db	661	ACTGCCCATCTGT	720
Qy	726	TGCAAGACTCAAGAAAACCTGACCTGCCAGGTGGCTACTCTCAACAAATGCCCTCACTG	785
Db	721	TGCAAGACTCAAGAAAACCTGACCTGCCAGGTGGCTACTCTCAACAAATGCCCTCACTG	780
Qy	786	AAGGACCTGT	845
Db	781	AAGGACCTGT	840
Qy	846	ACTCTGGGATGTCTTGGGCCGAGGCTGAGAAAGTACTGCCAGCTGAAGAACCCACCTGG	905
Db	841	ACTCTGGGATGTCTTGGGCCGAGGCTGAGAAAGTACTGCCAGCTGAAGAACCCACCTGG	900
Qy	906	TGTCATCAACTCCAGGAGGAGCAGAAATTTTGTCCAGAAATATCTAGGCTCCGCATACA	965
Db	901	TGTCATCAACTCCAGGAGGAGCAGAAATTTTGTCCAGAAATATCTAGGCTCCGCATACA	960
Qy	966	CCTGGATGGGCTCAGTGACCTGAGGAGCCTGGAAGTGGTGGATGGAGGAGACTATG	1025
Db	961	CCTGGATGGGCTCAGTGACCTGAGGAGCCTGGAAGTGGTGGATGGAGGAGACTATG	1020
Qy	1026	CGACCGGCTTCAGAACTGGAGGACGAGCCAGCAGCTGGCAGGGGACCGGGCTGG	1085
Db	1021	CGACCGGCTTCAGAACTGGAGGACGAGCCAGCAGCTGGCAGGGGACCGGGCTGG	1080
Qy	1086	GTGAGGCGAGGACTGTGTCTCACTTCCATCCAGACGGCAGGTGGAAATGACGACGTCTGC	1145
Db	1081	GTGAGGCGAGGACTGTGTCTCACTTCCATCCAGACGGCAGGTGGAAATGACGACGTCTGC	1140
Qy	1146	AGAGCCCTACACTGGGTCTGCGAGGCTGGGCTGGGTCAGACCCAGGAGAGTCACT	1205
Db	1141	AGAGCCCTACACTGGGTCTGCGAGGCTGGGCTGGGTCAGACCCAGGAGAGTCACT	1200
Qy	1206	GAGCTGCTTTGGTGGGACCAACCGGCCACAGAAATGGCGGTGGGAGGAGACTCTTCTC	1265
Db	1201	GAGCTGCTTTGGTGGGACCAACCGGCCACAGAAATGGCGGTGGGAGGAGACTCTTCTC	1260
Qy	1266	ACGACCTCTCCAGAACCGCTCTGGGAGAGAAATAAGCACTGGGAGATTGGAGCACTG	1325
Db	1261	ACGACCTCTCCAGAACCGCTCTGGGAGAGAAATAAGCACTGGGAGATTGGAGCACTG	1320
Qy	1326	CTAACATTTTGAATTTTTTCTCTTAAATTTTAAAGATGGTATAGTGTCTTAAAGCTT	1385
Db	1321	CTAACATTTTGAATTTTTTCTCTTAAATTTTAAAGATGGTATAGTGTCTTAAAGCTT	1380
Qy	1386	TTATTTTTTTTCAACTTTTGAAAGTCAACTTCAATGAAGGTATAATTTTTTACATAATAA	1445
Db	1381	TTATTTTTTTTCAACTTTTGAAAGTCAACTTCAATGAAGGTATAATTTTTTACATAATAA	1440
Qy	1446	AATGCACTCATTTT 1458 	

Db 1286 AACTTTTGAAGTCAACTTCATGAAGGTATAATTTTACATATAATAAAATGCACCTCATTT 1345

RESULT 6
AR561601
LOCUS
DEFINITION Sequence 9 from patent US 6756478. DNA linear PAT 08-OCT-2004
ACCESSION AR561601
VERSION AR561601.1 GI:53974520
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.

REFERENCE 1 (bases 1 to 1370)
AUTHORS Valladeau, J., Ravel, O., Bates, E.E.M., Ford, J., Saeland, S. and Lebecque, S.J.E.
TITLE Isolated mammalian membrane protein genes; related reagents
JOURNAL Patent: US 6756478-A 9 29-JUN-2004;
FEATURES
source Location/Qualifiers
1..1370
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN
Query Match 76.0%; Score 1108; DB 6; Length 1370;
Best Local Similarity 89.4%; Pred. No. 7.9e-252;
Matches 1288; Conservative 0; Mismatches 5; Indels 147; Gaps 3;

Qy 28 TAGGGCTCTCGGATTTTCAGACCCAGACAGCAGGAGCTCCAGTCACTCTACCCCAGCT 87
Db 44 TACTGTCCCTGGGATTTTCAGATCCAGACCCAGCAGGACTCCAGTCACTCTACCCCAGCT 103
Qy 88 CTCAGAGACACAGCGCTCCAACTCTGAGTGAGCTCCCACTCTGTGCTTGTGAGCAGCAA 147
Db 104 CTCAGAGACACAGCGCTCCAACTCTGAGTGAGCTCCCACTCTGTGCTTGTGAGCAGCAA 163
Qy 148 CCAACGTGGGAATCACACCTCCAGACCTCCAGACCTCCAGACCTCCAGACCTGGGCGCGGC 207
Db 164 CCAACGTGGGAATCACACCTCCAGACCTCCAGACCTCCAGACCTGGGCGCGGC 223
Qy 208 CTGCGCTCCATTTTCAGCTGTGACAACTTCAGAGCCGTGTGGCCCAAGCATGACAAGGAC 267
Db 224 CTGCGCTCCATTTTCAGCTGTGACAACTTCAGAGCCGTGTGGCCCAAGCATGACAAGGAC 283
Qy 268 GTATGAAAACCTTCAGTACTTGGAGATAAGGTGAAAGTCCAGGGTTTAAAAATGGGCC 327
Db 284 GTATGAAAACCTTCAGTACTTGGAGATAAGGTGAAAGTCCAGGGTTTAAAAATGGGCC 343
Qy 328 ACTTCCTCTCAGTCCCTCTCGAGCGTCTCCGCTCTGGGCCCTGCCATCTCCTGCTGTC 387
Db 344 ACTTCCTCTCAGTCCCTCTCGAGCGTCTCCGCTCTGGGCCCTGCCATCTCCTGCTGTC 362
Qy 388 CTGGGGCTCGGCTCGTCTGCTGCTGCTCATCATCTGTGTGGTGGATTCCAAAATTCCAA 447
Db 363 -----CTGCTGTGGTTCATCATCTGTGTGGTGGATTCCAAAATTCCAA 406
Qy 448 ATTTTCAGAGGACCTGTGACCTTCAGAGACAGATTTTTCAGAACTTCACTTCAACACATGT 507
Db 407 ATTTTCAGAGGACCTGTGACCTTCAGAGACAGATTTTTCAGAACTTCACTTCAACACATGT 466
Qy 508 GCGGGAGATCCAGGACCTGACTTCCAGGGCAGCAGCTTGGNAGAAACGATAGCATCTCT 567
Db 467 GCGGGAGATCCAGGACCTGACTTCCAGGGCAGCAGCTTGGNAGAAACGATAGCATCTCT 526
Qy 568 GAAAGCTGAGTGGAGGGTTTTCAGAGGAAACGCGCAGGAGGGGTATCTGAGCTCCAGGA 627
Db 527 GAAAGCTGAGTGGAGGGTTTTCAGAGGAAACGCGCAGG----- 564
Qy 628 ACACACTACGAGGAAGGCACACCTTAGGCCACTGTGCCCACTGTGTGTGTGTGCC 687
Db 565 -----C 565
Qy 688 AGTTTCATTTGAAATGCTCTCTCGAGTCCAGCAGCTGGTGCAGACCTGGAAGAACTGAC 747

Db 566 AGTTTCATTTCTGAATGCTCTCTGGAGTCCAGCAGCTGGTGCAGAGCTGGAAGAACTGAC 625
Qy 748 CTGCCAGGTGGCTACTCTCAACAACAAT-----GCCTCCACTGAAGGGACCTGCTG 798
Db 626 CTGCCAGGTGGCTACTCTCAACAACAATGGTGAAGAGCCCTCCACTGAAGGGACCTGCTG 685
Qy 799 CCCGCTCAACTGGGTGGAGCACAAGCAGCTGCTACTGTTCTCTCACTCTGGGATGTC 858
Db 686 CCCGCTCAACTGGGTGGAGCACAAGCAGCTGCTACTGTTCTCTCACTCTGGGATGTC 745
Qy 859 CTGGGCGGAGGCTGAGAAGTACTGCGAGCTGGAAGACGCCACCTGGTGGTCACTCAACTC 918
Db 746 CTGGGCGGAGGCTGAGAGTACTGCGAGCTGAAGACGCCACCTGGTGGTCACTCAACTC 805
Qy 919 CAGGGAGGAGCAGAAATTTTGTCCAGAAATATCTAGGCTCCGCATACACCTGGATGGCCT 978
Db 806 CAGGGAGGAGCAGAAATTTTGTCCAGAAATATCTAGGCTCCGCATACACCTGGATGGCCT 865
Qy 979 CAGTGACCTTGAAGGAGCTTGGAGTGGGATGGAACAGACTATGCCAGCCGCTTCCA 1038
Db 866 CAGTGACCTTGAAGGAGCTTGGAGTGGGATGGAACAGACTATGCCAGCCGCTTCCA 925
Qy 1039 GAACTGGAAGCCAGGCGCAGCAGACTGGCAGGGGCACGGCTGGGTGGAGCGAGGA 1098
Db 926 GAACTGGAAGCCAGGCGCAGCAGACTGGCAGGGGCACGGCTGGGTGGAGCGAGGA 985
Qy 1099 CTGTGCTCACTTCCATCCAGAGCGCAGGTGGAATGACGAGTCTGCCAGAGGCCCTACCA 1158
Db 986 CTGTGCTCACTTCCATCCAGAGCGCAGGTGGAATGACGAGTCTGCCAGAGGCCCTACCA 1045
Qy 1159 CTGGGTCTGGAGGCTGGCTGGTCCAGCAGCCAGGAGAGTCACTGAGCTGCTTTGG 1218
Db 1046 CTGGGTCTGGAGGCTGGCTGGTCCAGCAGCCAGGAGAGTCACTGAGCTGCTTTGG 1105
Qy 1219 TGGGACCACTCCGCGCCACAGAAATGGCGGTGGGAGGAGGACTCTTCTCAGCAGCTCTCGC 1278
Db 1106 TGGGACCACTCCGCGCCACAGAAATGGCGGTGGGAGGAGGACTCTTCTCAGCAGCTCTCGC 1165
Qy 1279 AAGACCGCTCTGGGAGAGAAATAGCACTGGGAGATGGAGCAGTCTGCTAACTTTTGA 1338
Db 1166 AAGACCGCTCTGGGAGAGAAATAGCACTGGGAGATGGAGCAGTCTGCTAACTTTTGA 1225
Qy 1339 TTTTCTCTTTTAAATTTTAAAGATGTTATAGTGTCTTAAGCTTTTATTTTTC 1398
Db 1226 TTTTCTCTTTTAAATTTTAAAGATGTTATAGTGTCTTAAGCTTTTATTTTTC 1285
Qy 1399 AACTTTTGAAGTCAACTTCATGAAGTATAATTTTACATAATAAAATGCACCTATT 1458
Db 1286 AACTTTTGAAGTCAACTTCATGAAGTATAATTTTACATAATAAAATGCACCTATT 1345

RESULT 7

BC027858

LOCUS

DEFINITION

Homo sapiens C-type (calcium dependent, carbohydrate-recognition

domain) lectin, superfamily member 14 (macrophage-derived), mRNA

(CDNA clone MGC:34439 IMAGE:5225381), complete cds.

ACCESSION

BC027858

VERSION

BC027858.1 GI:45708677

KEYWORDS

MGC.

SOURCE

Homo sapiens

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 1858)

AUTHORS

Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,

Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,

Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,

Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,

Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,

Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,

Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,

1858 bp mRNA linear PRI 06-APR-2004

Homo sapiens C-type (calcium dependent, carbohydrate-recognition

domain) lectin, superfamily member 14 (macrophage-derived), mRNA

(CDNA clone MGC:34439 IMAGE:5225381), complete cds.

BC027858

BC027858.1 GI:45708677

MGC.

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 1858)

Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,

Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,

Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,

Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,

Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,

Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,

Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,

Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strass, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (24-JAN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 176813)

REFERENCE
AUTHORS

Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,
Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B.,
Cammarata, J., Chang, J., Chazaro, B., Choepel, Y., Collamore, A.,
Cook, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J. S., Dodge, S.,
Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J.,
Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B.,
Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A.,
Karatas, A., Kells, C., Lander, E., Linton, L., Lindblad-Toh, K.,
Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C.,
McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mienga, V.,
Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H.,
O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K.,
Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Roy, A., Schauer, S., Schupback, R., Seaman, S., Severy, P.,
Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J.,
Testaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,
Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (27-MAR-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

On Sep 13, 2002 this sequence version replaced gi:22475358.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L22894

Center clone name: 467_D_22

* NOTE: This is a 'working draft' sequence. It currently

* consists of 8 contigs. Gaps between the contigs

* are represented as runs of N. The order of the pieces

* is believed to be correct as given, however the sizes

* of the gaps between them are based on estimates that have

* provided by the submitter.

* This sequence will be replaced

* by the finished sequence as soon as it is available and

* the accession number will be preserved.

* 947 1046: contig of 946 bp in length

* 1047 8054: contig of 7008 bp in length

* 8055 8154: gap of 100 bp

* 8155 50293: contig of 42139 bp in length

* 50294 50393: gap of 100 bp

* 50394 75289: contig of 24896 bp in length

* 75290 75389: gap of 100 bp

* 75390 98300: contig of 22911 bp in length

* 98301 98400: gap of 100 bp

* 98401 111897: contig of 13497 bp in length

* 111898 111997: gap of 100 bp

* 111998 144746: contig of 32749 bp in length

* 144747 144846: gap of 100 bp

* 144847 176813: contig of 31967 bp in length.

Location/Qualifiers

1. .176813

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosomes="17"

FEATURES
source

/map="17"
/clone=RP11-467D22"
/clone_lib="RPC1-11 Human Male BAC"

ORIGIN

Query Match	28.9%	Score 420.8;	DB 2;	Length 176813;
Best Local Similarity	99.5%	Pred. No. 1.7e-88;		
Matches 422;	Conservative 0;	Mismatches 2;	Indels 0;	Gaps 0;
QY 1035	TCCAGAACTGGAAGCCAGGCGCAGACGACTGCGAGGCGCAGCGGCTGGGTGGAGCGG	1094		
DB 149559	TCAGGAACCTGGAAGCCAGGCGCAGACGACTGCGAGGCGCAGCGGCTGGGTGGAGCGG	149618		
QY 1095	AGGACTGTGCTCACCTTCCATCCAGACGCGAGGTGAATGACGACGCTCTGCCAGAGGCCCT	1154		
DB 149619	AGGACTGTGCTCACCTTCCATCCAGACGCGAGGTGAATGACGACGCTCTGCCAGAGGCCCT	149678		
QY 1155	ACCACTGGGTCTGCGAGGCTGGCTGGGTCTAGACGAGCAGGAGAGTCACTGAGCTGCCT	1214		
DB 149679	ACCACTGGGTCTGCGAGGCTGGCTGGGTCTAGACGAGCAGGAGAGTCACTGAGCTGCCT	149738		
QY 1215	TTGGTGGGACCAACCCGCGCCACAGAAATGCGCGTGGGAGGAGGACTCTTCTCAGGACCTCC	1274		
DB 149739	TTGGTGGGACCAACCCGCGCCACAGAAATGCGCGTGGGAGGAGGACTCTTCTCAGGACCTCC	149798		
QY 1275	TGCAAGAACCGCTCTGGGAGAGAAATAGACACTGGGAGATTGGAAGCACTGCTAACATTT	1334		
DB 149799	TGCAAGAACCGCTCTGGGAGAGAAATAGACACTGGGAGATTGGAAGCACTGCTAACATTT	149858		
QY 1335	TGAATTTTTTCTCTTTAAATTTTAAAGATGGTATAGTGTCTTAAAGCTTTTATTTT	1394		
DB 149859	TGAATTTTTTCTCTTTAAATTTTAAAGATGGTATAGTGTCTTAAAGCTTTTATTTT	149918		
QY 1395	TTCCAACCTTTGAAAGTCAACTTCATGAAGGTATATTTTACATAATAAAAATGCATC	1454		
DB 149919	TTCCAACCTTTGAAAGTCAACTTCATGAAGGTATATTTTACATAATAAAAATGCATC	149978		
QY 1455	ATTT 1458			
DB 149979	ATTT 149982			

RESULT 12

AC120057/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 198821)

Birren, B., Nusbaum, C. and Lander, E.

Unpublished

2 (bases 1 to 198821)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,

Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L.,

Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,

Chazaro, B., Choepel, Y., Collamore, M., Collins, S., Collamore, A.,

Cook, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J. S., Dodge, S.,

Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,

Ginde, S., Gord, S., Coyette, M., Graham, L., Grand-Pierre, N.,

Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,

Kamat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R.,

Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G.,

MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,

McCarthy, M., McEwan, P., McKernan, K., Meldrim, J., Meneus, L.,

Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,

Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neil, D.,

Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,

Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupack, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (02-MAY-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 198821)

REFERENCE

AUTHORS

Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B., Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Fero, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schupack, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (16-JAN-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

4 (bases 1 to 198821)

REFERENCE

AUTHORS

Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, S., Arachchi, H. M., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., Dearellano, K., Diaz, J. S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Fero, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Maclean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupack, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V. S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (01-APR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

5 (bases 1 to 198821)

REFERENCE

AUTHORS

Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, S., Arachchi, H. M., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., Dearellano, K., Diaz, J. S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Fero, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Maclean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupack, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V. S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (30-APR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

5 (bases 1 to 198821)

REFERENCE

AUTHORS

Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, S., Arachchi, H. M., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., Dearellano, K., Diaz, J. S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Fero, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Maclean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupack, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V. S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (30-APR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

5 (bases 1 to 198821)

REFERENCE

AUTHORS

Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, S., Arachchi, H. M., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., Dearellano, K., Diaz, J. S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Fero, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Maclean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupack, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V. S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (30-APR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

5 (bases 1 to 198821)

REFERENCE

AUTHORS

Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, S., Arachchi, H. M., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., Dearellano, K., Diaz, J. S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Fero, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Maclean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupack, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V. S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (30-APR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

5 (bases 1 to 198821)

REFERENCE

AUTHORS

Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, S., Arachchi, H. M., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., Dearellano, K., Diaz, J. S., Dodge, S., Dooley, K.,

```
repeat_region complement(10678. .10801)
/rpt_family="FLAM_C"
repeat_region 10803. .10870
/rpt_family="MLT1A0"
repeat_region 10889. .10948
/rpt_family="MIR"
repeat_region complement(10949. .11246)
/rpt_family="AluY"
repeat_region 11247. .11291
/rpt_family="MIR"
repeat_region complement(11305. .11615)
/rpt_family="AluSg"
repeat_region complement(11616. .11749)
/rpt_family="AluJo"
repeat_region complement(11750. .11912)
/rpt_family="MLT1A0"
repeat_region complement(11962. .12319)
/rpt_family="MIR"
repeat_region complement(12434. .12736)
/rpt_family="AluSg"
repeat_region 13661. .13969
/rpt_family="AluY"
repeat_region complement(13996. .14307)
/rpt_family="AluSg"
repeat_region complement(15041. .15384)

Query Match 28.9%; Score 420.8; DB 9; Length 198821;
Best Local Similarity 99.5%; Pred. No. 1.7e-88;
Matches 422; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1035 TCAGAACTGGAGCCAGCCAGACGACTGGCAGGGGCA CGGGCTGGGTGGAGGCG 1094
DB 29372 TCAGAACTGGAGCCAGCCAGCCAGACGACTGGCAGGGGCA CGGGCTGGGTGGAGGCG 29313

QY 1095 AGCACTGTGCTCACTTCATCCAGACGGCAGGTGGAATCAGACGCTGCCAGAGCCCT 1154
DB 29312 AGCACTGTGCTCACTTCATCCAGACGGCAGGTGGAATCAGACGCTGCCAGAGCCCT 29253

QY 1155 ACCACTGGGTCTCGAGGCTGGCTGGGTTCAGACGAGGAGAGTCACTGAGCTGCCT 1214
DB 29252 ACCACTGGGTCTCGAGGCTGGCTGGGTTCAGACGAGGAGAGTCACTGAGCTGCCT 29193

QY 1215 TTGGTGGGACCCAGCCGACAGAAATGGCGTGGAGGAGGACTCTTCTCAGACCTCC 1274
DB 29192 TTGGTGGGACCCAGCCGACAGAAATGGCGTGGAGGAGGACTCTTCTCAGACCTCC 29133

QY 1275 TCAGACAGCGCTCTGGGAGAGAAATAGCACTGGGAGATTGGAAGCACTGCTAACATTT 1334
DB 29132 TCAGACAGCGCTCTGGGAGAGAAATAGCACTGGGAGATTGGAAGCACTGCTAACATTT 29073

QY 1335 TGAATTTTTTCTCTTTAAATTTTAAAAAGATGATAGTGTCTTAAGCTTTTATTTTTT 1394
DB 29072 TGAATTTTTTCTCTTTAAATTTTAAAAAGATGATAGTGTCTTAAGCTTTTATTTTTT 29013

QY 1395 TTCCAACTTTTGAAGTCAACTTTCATGAAGGTATAATTTTTTATCATATAATAAATGCACATC 1454
DB 29012 TTCCAACTTTTGAAGTCAACTTTCATGAAGGTATAATTTTTTATCATATAATAAATGCACATC 28953

QY 1455 ATTT 1458
DB 28952 ATTT 28949
```

RESULT 13

AC015918/c

LOCUS AC015918

DEFINITION Homo sapiens chromosome 17 clone CTD-316508 map 17, 21 unordered pieces.

ACCESSION AC015918

VERSION AC015918.10 GI:25140142

KEYWORDS HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_CANCELLED.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 220581)
Birren, B., Nusbaum, C. and Lander, E.
Homo sapiens chromosome 17, clone CTD-316508
Unpublished
2 (bases 1 to 220581)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M., Baldwin, J., Barna, N., Beckerly, R., Boguslavsky, L., Boukhgalter, B., Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Dearrellano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M., Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D., Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Lehoczy, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrim, J., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X., Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 220581)
Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B., Canarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Dearrellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kanat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Mathews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Milenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Riese, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (27-MAR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Nov 20, 2002 this sequence version replaced gi:25046454.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L851
Center clone name: 3165_O_8

* NOTE: This is a 'working draft' sequence. It currently consists of 21 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will be preserved.

* 1 5786: contig of 5786 bp in length
* 5787 5886: gap of 100 bp
* 5887 9996: contig of 4110 bp in length
* 9997 10096: gap of 100 bp
* 10097 12904: contig of 2808 bp in length
* 12905 13004: gap of 100 bp

TITLE
JOURNAL
COMMENT

```
* 13005 21890: contig of 8886 bp in length
* 21891 21990: gap of 100 bp
* 21991 26114: contig of 4124 bp in length
* 26115 26214: gap of 100 bp
* 26215 31344: contig of 5130 bp in length
* 31345 31444: gap of 100 bp
* 31445 48271: contig of 16827 bp in length
* 48272 48371: gap of 100 bp
* 48372 64150: contig of 15785 bp in length
* 64157 64256: gap of 100 bp
* 64257 68444: contig of 4188 bp in length
* 68445 75763: contig of 7219 bp in length
* 75764 75864: gap of 100 bp
* 75864 82941: contig of 7078 bp in length
* 82942 83041: gap of 100 bp
* 83042 116398: contig of 33357 bp in length
* 116399 116498: gap of 100 bp
* 116499 129682: contig of 13184 bp in length
* 129683 129782: gap of 100 bp
* 129783 136437: contig of 6655 bp in length
* 136438 136537: gap of 100 bp
* 136538 159184: contig of 22647 bp in length
* 159185 159285: gap of 100 bp
* 159286 163761: contig of 4477 bp in length
* 163762 163861: gap of 100 bp
* 163862 175449: contig of 11588 bp in length
* 175450 175549: gap of 100 bp
* 175550 187324: contig of 11775 bp in length
* 187325 187424: gap of 100 bp
* 187425 194425: contig of 7033 bp in length
* 194426 194557: gap of 100 bp
* 194558 197618: contig of 3061 bp in length
* 197619 197719: gap of 100 bp
* 197719 220581: contig of 22863 bp in length.

FEATURES
    source
        1. 220581
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosome="17"
            /map="17"
            /clone="CTD-316508"
            /clone_lib="CITD2 Human BAC"

ORIGIN
    Query Match      28.9%; Score 420.8; DB 2; Length 220581;
    Best Local Similarity 99.5%; Pred. No. 1.7e-88;
    Matches 422; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1035 TCAGAACTGGAGCCAGCCAGCCAGCACTGGCAGGCGCAGGCGCTGGGTGGAGGCG 1094
Db 98486 TCAGAACTGGAGCCAGCCAGCCAGCCAGCACTGGCAGGCGCAGGCGCTGGGTGGAGGCG 98427

Qy 1095 AGACTGTGCTCACTTCATCCAGACGGCAGGTGGAAATGAGGAGTGTGCGCAGAGGCCCT 1154
Db 98426 AGACTGTGCTCACTTCATCCAGACGGCAGGTGGAAATGAGGAGTGTGCGCAGAGGCCCT 98367

Qy 1155 ACCACTGGGTCTGGAGCTGGCTGGCTCAGACCCAGCAGGAGAGTCACTGAGCTGCCT 1214
Db 98366 ACCACTGGGTCTGGAGCTGGCTGGCTCAGACCCAGCAGGAGAGTCACTGAGCTGCCT 98307

Qy 1215 TTGGTGGGACACCCGCGCCACAGAAATGGCGGTGGGAGGAGGACTCTTCTCAGGACCTCC 1274
Db 98306 TTGGTGGGACACCCGCGCCACAGAAATGGCGGTGGGAGGAGGACTCTTCTCAGGACCTCC 98247

Qy 1275 TCGAAGACCGCTCTGGGAGAGAAATAGCACTGGGAGATTTGGAGCACTGCTTAA CATT 1334
Db 98246 TCGAAGACCGCTCTGGGAGAGAAATAGCACTGGGAGATTTGGAGCACTGCTTAA CATT 98187

Qy 1335 TGAATTTTTTCTCTTAAATTTTAAAGATGGTATAGTCTTCTTAAGCTTTTATTTTTT 1394
Db 98186 TGAATTTTTTCTCTTAAATTTTAAAGATGGTATAGTCTTCTTAAGCTTTTATTTTTT 98127
```

```
Qy 1395 TTCCAACCTTTTGAAGTCAACTTCATGAGGTATATATTTTACATATAAAATGCACTC 1454
Db 98126 TTCCAACCTTTTGAAGTCAACTTCATGAGGTATATATTTTACATATAAAATGCACTC 98067

Qy 1455 ATTT 1458
Db 98066 ATTT 98063

RESULT 14
S36676
LOCUS
DEFINITION
    galactose and N-acetylgalactosamine-specific lectin [mice,
    S36676
    S36676.1 GI:249360
SOURCE
    Mus sp.
ORGANISM
    Mus sp.
    Rukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
    1 (bases 1 to 1414)
AUTHORS
    Sato, M., Kawakami, K., Osawa, T. and Toyoshima, S.
TITLE
    Molecular cloning and expression of cDNA encoding a
    galactose/N-acetylgalactosamine-specific lectin on mouse
    tumoricidal macrophages
JOURNAL
    J. Biochem. 111 (3), 331-336 (1992)
MEDLINE
    92268032
PUBMED
    1587794
REMARK
    GenBank staff at the National Library of Medicine created this
    entry [NCBI gisbseq 104199] from the original journal article.
FEATURES
    Location/Qualifiers
        1. 1414
            /organism="Mus sp."
            /mol_type="mRNA"
            /db_xref="taxon:10095"
        1. 1414
            /gene="galactose and N-acetylgalactosamine-specific
            lectin, MWGL"
        171..1085
            /gene="galactose and N-acetylgalactosamine-specific
            lectin, MWGL"
            /note="Method: conceptual translation with partial peptide
            sequencing; MWGL"
            /codon_start=1
            /product="galactose and N-acetylgalactosamine-specific
            lectin"
            /protein_id="AAB22171.1"
            /db_xref="GI:249361"
            /translation="MIYENLQNSRIEKTQEPKAPSQSFLWRILSWTHLLPSLGLS
            LLLVSVVIGSQSLRRDLGTLRATLNDTTSKIRAEQSLDSRADSPFKGISLKV
            DVEDHROELQAGRDLSQKVTLSLSTVEKREQALTDLSLTDHVOQLRKDLKALTCOL
            ANLKNGSVACCPDLHWTHEGSCYWFSEKSPKADKYCRLENSHLVVNSLEON
            FLQRLANVVSVIGLTDQNPWRVVDGTFPEKFKWAPLOPNWFGHGLGGEDCAH
            ITTGPWDDVDCQRTFRWICEMKLAKES"

ORIGIN
    Query Match      27.8%; Score 404.6; DB 10; Length 1414;
    Best Local Similarity 66.4%; Pred. No. 6.8e-85;
    Matches 635; Conservative 0; Mismatches 304; Indels 18; Gaps 3;

Qy 242 CGTGTGGCCCAAGCATGACAGGAGCTATGAAACTTCACGACTTTGGAGAAATAGGTG 301
Db 150 CGTGTGTGTCAAGTGTAAATGATATACGAAACCTCCAGAACTCAAGGATCGAGGAG 209

Qy 302 AAAGTCCAGGGGTTTAAATATGGGCCACTTCTCTCCAGTCCCTCTGCGAGCGTCTCCGC 361
Db 210 AAAACCAAGAGCGCTGTGAAGC-----TCCTTCCAGTCTCTTCTGTGCGCATCTC 263

Qy 362 TCTGGGCCCTGCCATCTCTCTGCTCTCCCTGGGCGCTGGCGCTGCTGCTGCTCATATC 421
Db 264 TCTGGACCCACCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 323
```


This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 27, 2005, 19:09:15 ; Search time 838 Seconds
(without alignments)
10299.502 Million cell updates/sec

Title: US-10-829-107-3

Perfect score: 1458

Sequence: 1 gttgagagatggatgtcc.....taataaaatgcactcattt 1458

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_16Dec04:*
1: Geneseq1980s:*
2: Geneseq1990s:*
3: Geneseq2000s:*
4: Geneseq2001as:*
5: Geneseq2001bs:*
6: Geneseq2002as:*
7: Geneseq2002bs:*
8: Geneseq2003as:*
9: Geneseq2003bs:*
10: Geneseq2003cs:*
11: Geneseq2003ds:*
12: Geneseq2004as:*
13: Geneseq2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1458	100.0	1458	2	AAx04866 Primate D
2	1434.2	98.4	1531	6	Abk94927 Human nov
3	1108	76.0	1370	2	Aax04868 Variant p
4	1039.2	71.3	1072	8	AcA10155 Human NOV
5	724.2	49.7	929	3	Aaf17999 Lung canc
6	717.6	49.2	944	12	Ado08342 Human NOV
7	714.6	49.0	820	8	AcA10154 Human NOV
8	714.6	49.0	820	12	Ado08340 Human NOV
9	606.6	41.6	661	4	Aaf44889 Human bre
10	438	30.0	1358	10	Abt42075 Toxicity
11	434	29.8	1432	13	Acn41226 Human dia
12	431.4	29.6	1477	13	Acn41224 Human dia
13	382	26.2	1277	6	Abn95732 Gene #223
14	382	26.2	1277	10	AcA56588 Human sig
15	382	26.2	1277	12	Adi56384 Human pol
16	380.4	26.1	1277	2	Aat66950 Asialogly
17	380.4	26.1	1295	4	Aah57466 Human liv
18	336.6	23.1	855	10	AdB52510 Primary r
19	321.4	22.0	1055	2	AcD06173 Human CDN
20	321.4	22.0	1300	2	Aat66951 Asialogly

21	321.4	22.0	1300	6	ABN97236 Gene #373
22	313.4	21.5	1055	12	AdN95955 Human NOV
23	312.6	21.4	1276	13	Acn41228 Human dia
24	310.4	21.3	361	3	AAC09703 Human sec
25	307.6	21.1	1418	13	Acn41227 Human dia
26	305	20.9	1463	13	Acn41225 Human dia
27	303	20.8	1112	9	AcD06172 Human CDN
28	299.8	20.6	1112	12	AdN95953 Human NOV
29	282.8	19.4	1309	6	Abn95737 Gene #223
30	282.8	19.4	1309	9	AcD06171 Human CDN
31	282.8	19.4	1309	10	AdB4950 Farnesyl
32	282.8	19.4	1493	13	AdP23958 PRO polyp
33	267.2	18.3	1309	12	AdN95951 Human NOV
34	262.4	18.0	780	12	AdO21128 Human car
35	246	16.9	443	8	AbX41898 Bovine ES
36	233	16.0	1346	13	Acn41223 Human dia
37	233	16.0	1391	13	Acn41222 Human dia
38	231.8	15.9	466	9	ACH21572 Human adu
39	220	15.1	566	4	Aah57258 Human liv
40	212.4	14.6	1290	6	Abk63735 Rat seque
41	212.4	14.6	1290	10	AdB58277 Toxicity-
42	212.4	14.6	1290	10	AdB582814 Primary r
43	211	14.5	521	13	AdQ58274 Novel can
44	183	12.6	400	3	AAC57412 Arachidon
45	155.6	10.7	448	3	AAC57415 Arachidon

ALIGNMENTS

RESULT 1

AAx04866
ID AAX04866 standard; CDNA; 1458 BP.

XX
AC AAX04866;

DT 11-MAY-1999 (first entry)

XX Primate DCM2 C-lectin family gene nucleotide sequence.

DE ss; Primate; dendritic cell membrane protein; DCM2P1; DCM2P2;

KW chromosomal abnormality; expression misregulation;

KW abnormal proliferation; regeneration; degeneration; haematopoietic cell.

XX Mammalia.

PH Key Location/Qualifiers

FT CDS 257..1207

FT /*tag= a /product= "DCMP2 C-lectin family gene protein"

XX WO9902562-A1.

PN 21-JAN-1999.

XX 08-JUL-1998; 98WO-US013436.

XX 09-JUL-1997; 97US-0053080P.

XX (SCHE) SCHERING CORP.

XX Valladeau J, Ravel O, Bates EEM, Ford J, Saeland S, Lebecque SJE;

XX WPI; 1999-120786/10.

XX Dendritic cell membrane proteins - used to treat conditions associated with abnormal physiology or development.

XX Disclosure; Page 68-69; 82pp; English.

XX Dendritic cell membrane protein 1 (DCMP1) and DCM2P2 nucleic acids can be used as markers for distinguishing cell types, including genomic aspects of cells, as well as mRNA and protein expression patterns. They can also

CC be used to detect chromosomal abnormalities. The proteins can be used to
CC diagnose disorders associated with expression misregulation. They can
CC also be used to treat conditions associated with abnormal physiology or
CC development, including abnormal proliferation, e.g. cancerous conditions
CC or degenerative conditions. Abnormal proliferation, regeneration, the
CC degeneration and atrophy may be modulated using the proteins. The
CC proteins may also play a role in regulation or development of
CC haematopoietic cells
XX
SQ

Sequence 1458 BP; 349 A; 406 C; 388 G; 315 T; 0 U; 0 Other;

Query Match 100.0%; Score 1458; DB 2; Length 1458;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1458; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	GTGTGAGGATGGGATGCTCCAGATCATAGGGCTCTGGGATTTGAGACCCAGACGAGC	60
DB	1	GTGTGAGGATGGGATGCTCCAGATCATAGGGCTCTGGGATTTGAGACCCAGACGAGC	60
QY	61	AGGACTCCAGTCACTTACCCAGCTCTCCAGGACACAGCGCTCCCAACTCTGAGTGAC	120
DB	61	AGGACTCCAGTCACTTACCCAGCTCTCCAGGACACAGCGCTCCCAACTCTGAGTGAC	120
QY	121	GTCCCACTCTGTGCTTGTGAGCACAAACAGTGGGAATCACACCTCCAGACCTCCCA	180
DB	121	GTCCCACTCTGTGCTTGTGAGCACAAACAGTGGGAATCACACCTCCAGACCTCCCA	180
QY	181	CAGCTCCACCCAGATGGGGCGGCGCTCTCCATTTTCACTGTGACCACTCAGAG	240
DB	181	CAGCTCCACCCAGATGGGGCGGCGCTCTCCATTTTCACTGTGACCACTCAGAG	240
QY	241	CCGTGTTGGCCCAAGCATGACAGGACGTATGAAACTTCCAGTACTTGGAGAAATAGGT	300
DB	241	CCGTGTTGGCCCAAGCATGACAGGACGTATGAAACTTCCAGTACTTGGAGAAATAGGT	300
QY	301	GAAAGTCCAGGGGTTTAAAAATGGGCCACTTCTCTCCAGTCCCTCTCGACGCTCCG	360
DB	301	GAAAGTCCAGGGGTTTAAAAATGGGCCACTTCTCTCCAGTCCCTCTCGACGCTCCG	360
QY	361	CTCTGGGCGCTGCATCTCTGCTGTCTCTGGGCTCTGGGCTCTGCTGTCTGCTCAT	420
DB	361	CTCTGGGCGCTGCATCTCTGCTGTCTCTGGGCTCTGGGCTCTGCTGTCTGCTCAT	420
QY	421	CTGTGTGTTGGATTCCAAATTTCCAGAGGACCTGGTGACCTCAGAACACAGA	480
DB	421	CTGTGTGTTGGATTCCAAATTTCCAAATTTCCAGAGGACCTGGTGACCTCAGAACACAGA	480
QY	481	TTTTAGCAACTTCACTCAAAACACTGTGGCGGAGATCCAGGCACTGACTTCCAGGGCAG	540
DB	481	TTTTAGCAACTTCACTCAAAACACTGTGGCGGAGATCCAGGCACTGACTTCCAGGGCAG	540
QY	541	CAGCTTGGAGAACATAGCATCTCTGAAAGCTGAGGTGGAGGTTTCAAGCAGGAACG	600
DB	541	CAGCTTGGAGAACATAGCATCTCTGAAAGCTGAGGTGGAGGTTTCAAGCAGGAACG	600
QY	601	GCAGGACGGGTTATCTGAGCTCCAGGAACACACTACGACAGAGGACACCTAGGCCACTG	660
DB	601	GCAGGACGGGTTATCTGAGCTCCAGGAACACACTACGACAGAGGACACCTAGGCCACTG	660
QY	661	TCCCACTGCCCATCTGTGTGTCCTCCAGTTCACTTCTGAAATGCTCTCTCGAGTCCAGCA	720
DB	661	TCCCACTGCCCATCTGTGTGTCCTCCAGTTCACTTCTGAAATGCTCTCTCGAGTCCAGCA	720
QY	721	GCTGGTGCAGACCTGAAGAACTGACCTGCCAGGTGGCTACTCTCAACAAATGCTCTC	780
DB	721	GCTGGTGCAGACCTGAAGAACTGACCTGCCAGGTGGCTACTCTCTCAACAAATGCTCTC	780
QY	781	CAGTGAAGGACCTGTGCTGCCCCCTCAACTGGGTGGAGACCAAGACAGCTGCTACTGTT	840
DB	781	CAGTGAAGGACCTGTGCTGCCCCCTCAACTGGGTGGAGACCAAGACAGCTGCTACTGTT	840
QY	841	CTCTCACTCTGGGATGCTCTGGGCGGAGGTGAGAAATGCTGCCAGCTGAAGAACGCCCA	900
DB	841	CTCTCACTCTGGGATGCTCTGGGCGGAGGTGAGAAATGCTGCCAGCTGAAGAACGCCCA	900

DB	841	CTCTCACTCTGGGATGCTCTGGGCGGAGGTGAGAAATGCTGCCAGCTGAAGAACGCCCA	900
QY	901	CCTGTGTCTCATCACTCCAGGGAGGACAGAAATTTTCTCCAGAAATATCTAGGCTCCGC	960
DB	901	CCTGTGTGTCTCATCACTCCAGGGAGGACAGAAATTTTCTCCAGAAATATCTAGGCTCCGC	960
QY	961	ATACACCTTGGATGGGCTTCAGTGACCCCTGAAGGACCTTGGATGGATGGAACAGA	1020
DB	961	ATACACCTTGGATGGGCTTCAGTGACCCCTGAAGGACCTTGGATGGATGGAACAGA	1020
QY	1021	CTATCGGACCGGCTTCCAGAACTGGAAGCCAGGCCAGCCAGACGACTGGCAGGGGACAG	1080
DB	1021	CTATCGGACCGGCTTCCAGAACTGGAAGCCAGGCCAGCCAGACGACTGGCAGGGGACAG	1080
QY	1081	GCTGGGTGGAGCGGAGGACTGTGCTCACTTCCATCCAGACGCGAGGTGGAATGACGAGT	1140
DB	1081	GCTGGGTGGAGCGGAGGACTGTGCTCACTTCCATCCAGACGCGAGGTGGAATGACGAGT	1140
QY	1141	CTGCCAGAGGCGCTTACCACCTGGGTCTGCGAGGCTGGCTGGGTCTGAGACCCAGGAGAG	1200
DB	1141	CTGCCAGAGGCGCTTACCACCTGGGTCTGCGAGGCTGGCTGGGTCTGAGACCCAGGAGAG	1200
QY	1201	TCACGTGAGCTGCTTGTGGGACCAACCCGGGACAGAAATGGCGTGGAGGAGGACTC	1260
DB	1201	TCACGTGAGCTGCTTGTGGGACCAACCCGGGACAGAAATGGCGTGGAGGAGGACTC	1260
QY	1261	TTCTCACGACCTCTCCGCAAGACCGCTCTGGGAGAGAAATAGCACTGGGAGATGGAG	1320
DB	1261	TTCTCACGACCTCTCCGCAAGACCGCTCTGGGAGAGAAATAGCACTGGGAGATGGAG	1320
QY	1321	CACCTCTAACATTTTGAATTTTCTCTTAAATTTTAAAGATGGTATAGTGTCTTA	1380
DB	1321	CACCTCTAACATTTTGAATTTTCTCTTAAATTTTAAAGATGGTATAGTGTCTTA	1380
QY	1381	AGCTTTTATTTTTCCTTCAACTTTTGAAGTCACTTCATGAAAGTAAATTTTACATA	1440
DB	1381	AGCTTTTATTTTTCCTTCAACTTTTGAAGTCACTTCATGAAAGTAAATTTTACATA	1440
QY	1441	ATAAAATGCACCTCATTT 1458	
DB	1441	ATAAAATGCACCTCATTT 1458	

RESULT 2
ABK94927
ID ABK94927 standard; cDNA; 1531 BP.
XX
AC ABK94927;
XX
DT 30-AUG-2002 (first entry)
XX
DE Human novel polynucleotide #38.
XX
KW Human; gene; ss; inflammatory condition; shock; sepsis; immune response;
KW cancer; wound healing; central nervous system disease; haematopoiesis;
KW peripheral nervous system disease; amyotrophic lateral sclerosis; tendon;
KW myeloid cell disorder; lymphoid cell disorder; platelet disorder; bone;
KW cartilage; ligament; nerve tissue; ulcer; osteoporosis; osteoarthritis;
KW bone degenerative disorder; periodontal disease; reperfusion injury;
KW lung fibrosis; liver fibrosis; autoimmune disorder; bacterial infection;
KW allergic condition; thrombolytic; thrombosis; coagulation disorder;
KW fungal infection.
XX
OS Homo sapiens.
XX
PN WO200244340-A2.
XX
PD 06-JUN-2002.
XX
PF 30-NOV-2001; 2001WO-US047004.
XX
PR 30-NOV-2000; 2000US-00728952.
XX

QY 1339 TTTTCTCTCTTAATTTTAAAGAGATGGTATAGTGTCTTAAGCTTTTATTTTTC 1398
 Db 1226 TTTTCTCTCTTAATTTTAAAGAGATGGTATAGTGTCTTAAGCTTTTATTTTTC 1285
 QY 1399 AACTTTTGAAGCAACTTCATGAAGGTATAATTTTACATAATAAATAATGCACTATT 1458
 Db 1286 AACTTTTGAAGCAACTTCATGAAGGTATAATTTTACATAATAAATAATGCACTATT 1345

RESULT 4

ACA10155
 ID ACA10155 standard; cDNA; 1072 BP.
 AC ACA10155;
 XX
 DT 02-JUN-2003 (first entry)
 XX
 DE Human NOVX polynucleotide #45.
 XX
 KW Human; NOVX; gene; ss; metabolic disorder; diabetes; infectious disease;
 KW obesity; anorexia; cancer; cardiovascular disorder; asthma; neurogenesis;
 KW neurodegenerative disorder; epilepsy; immune disorder; osteoarthritis;
 KW haematopoietic disorder; inflammatory skin disorder; dyslipidemia;
 KW haematopoiesis; wound healing; angiogenesis; bacterial infection;
 KW viral infection; fungal infection; helminthic infection; atherosclerosis;
 KW protozoal infection; hypertension.
 XX
 OS Homo sapiens.
 XX
 PN W0200290504-A2.
 XX
 PD 14-NOV-2002.

XX
 PF 02-MAY-2002; 2002WO-US014342.
 XX
 PR 03-MAY-2001; 2001US-0288395P.
 PR 04-MAY-2001; 2001US-0288900P.
 PR 07-MAY-2001; 2001US-0289087P.
 PR 14-MAY-2001; 2001US-0290753P.
 PR 15-MAY-2001; 2001US-0291189P.
 PR 16-MAY-2001; 2001US-0291243P.
 PR 18-MAY-2001; 2001US-0292001P.
 PR 21-MAY-2001; 2001US-0292374P.
 PR 22-MAY-2001; 2001US-0292587P.
 PR 23-MAY-2001; 2001US-0293107P.
 PR 28-MAY-2001; 2001US-0294110P.
 PR 30-MAY-2001; 2001US-0294434P.
 PR 31-MAY-2001; 2001US-0294827P.
 PR 18-JUN-2001; 2001US-0298988P.
 PR 31-JUL-2001; 2001US-0308901P.
 PR 17-AUG-2001; 2001US-0313388P.
 PR 21-AUG-2001; 2001US-0313851P.
 PR 21-AUG-2001; 2001US-0313937P.
 PR 17-SEP-2001; 2001US-0322701P.
 PR 17-SEP-2001; 2001US-0322802P.
 PR 25-SEP-2001; 2001US-0324757P.
 PR 27-SEP-2001; 2001US-0325314P.
 PR 27-SEP-2001; 2001US-0325682P.
 PR 21-NOV-2001; 2001US-0332129P.
 PR 03-DEC-2001; 2001US-0336882P.
 PR 14-DEC-2001; 2001US-0340305P.
 PR 01-MAY-2002; 2002US-00138588.

(CURA-) CURAGEN CORP.

XX
 PI Alsobrook JP, Anderson DM, Boldog FL, Burgess CE, Casman SJ;
 PI Chapoval A, Edinger S, Gerlach V, Gorman L, Gunther E, Guo X;
 PI Kekula R, Lepley DM, Li L, Liu X, Malyankar UM, Miller CE;
 PI Millet I, Padigaru M, Patturajan M, Pena CE, Rieger DK, Sheny SG;
 PI Shimkets RA, Spytek KA, Taupier RJ, Vernet CAM, Voss EZ;
 PI Zerhusen BD;

WPI; 2003-103512/09.

P-PSDB; ABUS9170.

XX
 PT New isolated NOVX polypeptides and polynucleotides, useful for
 PT preventing, diagnosing or treating NOVX-associated disorders, e.g.
 PT osteoarthritis, obesity, atherosclerosis, cancer, Parkinson's disease,
 PT asthma, or infections.

Claim 20; Page 169-170; 340pp; English.

XX
 CC The invention relates to human NOVX polypeptides and the polynucleotides
 CC encoding them. The polypeptides, polynucleotides and antibodies that bind
 CC immunospecifically to the polypeptides are useful in the manufacture of a
 CC medicament for treating a syndrome associated with a human disease,
 CC preferably a NOVX-associated disorder. The sequences are useful for
 CC treating, preventing or diagnosing diseases such as metabolic disorders,
 CC diabetes, obesity, infectious diseases (viral, bacterial, fungal,
 CC helminthic, and protozoal), anorexia, cancer, cardiovascular disorders
 CC (e.g. hypertension, atherosclerosis), neurodegenerative disorders (e.g.
 CC Alzheimer's disease, Parkinson's disease), epilepsy, immune disorders,
 CC osteoarthritis, haematopoietic disorders, inflammatory skin disorders,
 CC asthma and various dyslipidemias. The nucleic acids and polypeptides may
 CC also be used as targets for the identification of small molecules that
 CC modulate or inhibit e.g. neurogenesis, cell differentiation, cell
 CC proliferation, haematopoiesis, wound healing and angiogenesis, and in the
 CC generation of antibodies that bind immunospecifically to NOVX substances
 CC for use in therapeutic or diagnostic methods. The nucleic acids are
 CC further used as hybridisation probes, and in chromosome mapping, tissue
 CC typing, preventive medicine and pharmacogenomics. Sequences ACA1011-
 CC ACA10156 represent human NOVX polynucleotides of the invention
 XX
 SQ Sequence 1072 BP; 247 A; 297 C; 313 G; 215 T; 0 U; 0 Other;

Query Match 71.3%; Score 1039.2; DB 8; Length 1072;
 Best Local Similarity 98.9%; Pred. No. 7.5e-259;
 Matches 1060; Conservative 0; Mismatches 3; Indels 9; Gaps 1;
 QY 213 CTCATTTTCAGCTGTGACACCTCAGAGCCGTGTGGCCCAAGCATCAGAGGAGTATG 272
 Db 1 CTCATTTTCAGCTGTGACACCTCAGAGCCGTGTGGCCCAAGCATCAGAGGAGTATG 60
 QY 273 AAAAATTCAGTACTTGGAGATAAGGTGAAGTCCAGGGGTATAAAATGGGCACTTC 332
 Db 61 AAAATTCAGTACTTGGAGATAAGGTGAAGTCCAGGGGTATAAAATGGGCACTTC 120
 QY 333 CTCTCCAGTCCCTCTCGAGCGTCTCGCTCTGGGCCCTGCCATCTCCTGCTCCCTGG 392
 Db 121 CTCTCCAGTCCCTCTCGAGCGTCTCGCTCTGGGCCCTGCCATCTCCTGCTCCCTGG 180
 QY 393 GCCTCGGCTGCTGCTGCTGCTCATCATCTGTGTGGTGGATTCCAAATTCGAATTC 452
 Db 181 GCCTCGGCTGCTGCTGCTGCTCATCATCTGTGTGGTGGATTCCAAATTCGAATTC 240
 QY 453 AGAGGGACCTGGTGACCCCTGAGAACAGATTTTGTAGCACTTCCACCTCAACACTCTGGCGG 512
 Db 241 AGAGGGACCTGGTGACCCCTGAGAACAGATTTTGTAGCACTTCCACCTCAACACTCTGGCGG 300
 QY 513 AGATCCAGGCACTGACTTCCAGGGCAGCAGCTTGGAGAAACGATAGCATCTCTGAAAG 572
 Db 301 AGATCCAGGCACTGACTTCCAGGGCAGCAGCTTGGAGAAACGATAGCATCTCTGAAAG 360
 QY 573 CTGAGTGGAGGGTTTCAACGAGGAACGGCAGGGGTATCTGAGCTCCAGGAACACA 632
 Db 361 CTGAGTGGAGGGTTTCAACGAGGAACGGCAGGGGTATCTGAGCTCCAGGAACACA 420
 QY 633 CTAGCGAAGAGGCACACCTAGGCCACTGTCCCACTGCCCATCTGTGTGTCTCCAGTTC 692
 Db 421 CTAGCGAAGAGGCACACCTAGGCCACTGTCCCACTGCCCATCTGTGTGTCTCCAGTTC 480
 QY 693 ATTCTGAATGCTCTCTGCGAGTCCAGCAGCTGGTGCAGAACCTGGAAGAACTGACCTGCC 752
 Db 481 ATTCTGAATGCTCTCTGCGAGTCCAGCAGCTGGTGCAGAACCTGGAAGAACTGACCTGCC 540
 QY 753 AGGTGGCTACTCTCAACACAAT-----GCCTCCACTGAAGGGACCTGTGCCCCG 803

Db 541 AGGTGGCTACTCTCAACAACAAATGGTGAAGAGCCTCCACTGAAGGGAGCCTGTCGCCCTG 600
QY 804 TCAACTGGGTGGAGCACCAGACAGCTGTACTGGTCTCTCACTCTGGAGTGTCTCTGG 863
Db 601 TCAACTGGGTGGAGCACCAGACAGCTGTACTGGTCTCTCACTCTGGAGTGTCTCTGG 660
QY 864 CCAGAGCTGAGAAAGTACTGCGAGCTGAAGAAACCCACCTGGTGTGTCATCACTCCAGGG 923
Db 661 CCGAGGCTGAGAAGTACTGCCAGCTGAAGAACGCCACCTGGTGTGTCATCACTCCAGGG 720
QY 924 AGGAGCAGAAATTTGTCGAGAAATATCTAGGCTCCGATACACCTGGATGGGCTTCAGTG 983
Db 721 AGGAGCAGAAATTTGTCGAGAAATATCTAGGCTCCGATACACCTGGATGGGCTTCAGTG 780
QY 984 ACCTTGAAGAGCCTGGAGTGGTGGATGGAGACAGACTATGCGACGGCTTCCAGAACT 1043
Db 781 ACCCTGAAGAGCCTGGAGTGGTGGATGGAGACAGACTATGCGACGGCTTCCAGAACT 840
QY 1044 GGAAGCCAGGCCAGCCAGACCACTGCGCAGGGGACCGGCTGGGTGGAGCGAGGACTGTG 1103
Db 841 GGAAGCCAGGCCAGCCAGACCACTGCGCAGGGGACCGGCTGGGTGGAGCGGAGGACTGTG 900
QY 1104 CTCACCTTCCATCCAGACGGCAGGTGGAATGACGACTGTGTCAGAGGCGCTTACCCTGGG 1163
Db 901 CTCACCTTCCATCCAGACGGCAGGTGGAATGACGACTGTGTCAGAGGCGCTTACCCTGGG 960
QY 1164 TCTGCGAGGCTGCGCTGGTTCAGACAGCCAGGAGAGTCACTGAGCTGCTTTGGTGGGA 1223
Db 961 TCTGCGAGGCTGCGCTGGTTCAGACAGCCAGGAGAGTCACTGAGCTGCTTTGGTGGGA 1020
QY 1224 CCACCCGGCCACAGAAATGGCGTGGGAGGAGGACTCTTCTCACGACCTCCT 1275
Db 1021 CCACCCGGCCACAGAAATGGCGTGGGAGGAGGACTCTTCTCACGACCTCCT 1072

RESULT 5
ID AAF17999 standard; DNA; 929 BP.
XX AAF17999;
AC AAF17999;
XX
DT 14-MAR-2001 (first entry)
XX Lung cancer associated polynucleotide sequence SEQ ID 18.
DE
DE Human; lung cancer associated protein; neuroprotective; cytostatic;
KW cardioactive; immunomodulatory; muscular active; vulnary;
KW gastrointestinal; nephrotropic; antiinfective; gynecological;
KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
KW proliferative disorder; wound healing; infectious disease; ds.
XX
OS Homo sapiens.
XX
XX WO200055180-A2.
PN
XX 21-SEP-2000.
XX
XX 08-MAR-2000; 2000WO-US005918.
PF
XX 12-MAR-1999; 99US-0124270P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA (ROSE/) ROSEN C A.
XX
XX Ruben SM;
PI
XX WPI; 2000-587514/55.
XX
XX P-PSDB; AAB58123.
DR
XX Lung cancer associated gene sequences, referred to as lung cancer
PT antigens, useful for treatment, prevention, and diagnosis of disorders
PT such as lung cancer.

XX
PS Claim 1; Page 503; 1425pp; English.
XX
CC Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
CC associated proteins represented in AAB58106 - AAB58548. Lung cancer
CC associated proteins and polynucleotide sequences, their agonists, and
CC antagonists may have neuroprotective; cytostatic; cardioactive;
CC immunomodulatory; muscular active general; vulnary; gastrointestinal
CC general; nephrotropic; antiinfective; gynecological; or antibacterial
CC activity. The invention also includes antibodies specific for the protein
CC or polynucleotide sequences. The lung cancer associated polynucleotide
CC sequences may be used for detection of lung cancer, chromosome
CC identification, as chromosome markers, and for numerous other diagnostic
CC or research purposes. The proteins may be used to treat disorders such as
CC neural, immune, muscular, reproductive, gastrointestinal, pulmonary,
CC cardiovascular, renal, and proliferative disorders. The proteins may also
CC be used in the treatment of wounds and infectious diseases.
CC Polynucleotide sequences AAF18425 - AAF18433 and peptide AAB58549 are
CC used in the course of the invention for the identification and
CC characterisation of the polynucleotide and protein sequences
XX
SQ Sequence 929 BP; 217 A; 267 C; 234 G; 201 T; 0 U; 10 Other;
Query Match 49.7%; Score 724.2; DB 3; Length 929;
Best Local Similarity 89.8%; Pred. No. 3.6e-177;
Matches 836; Conservative 6; Mismatches 6; Indels 83; Gaps 3;
QY 1 GTTGGAGATGGGATGCCAGATGATAGGCTCTCTGGGATTTCCAGACCCAGACCAG 60
Db 32 GTTGGAGATGGGATGCCAGATGATAGGCTCTCTGGGATTTCCAGACCCAGACCAG 91
QY 61 AGGACTCCAGTCACTCTACCCAGCTCTCCAGGACACAGCGCTCCCAACTCTGAGTGAC 120
Db 92 AGGACTCCAGTCACTCTACCCAGCTCTCCAGGACACAGCGCTCCCAACTCTGAGTGAC 151
QY 121 GTCCACCTCTGGTCTTTCAGACACAACTGGGAAATCAACCTCCAGACCTCCCA 180
Db 152 GTCCACCTCTGGTCTTTCAGACACAACTGGGAAATCAACCTCCAGACCTCCCA 211
QY 181 CAGCTCCACCCAGATGGGCGCGGCTGCTCCATTTCCAGCTGTGACACCTCAGAG 240
Db 212 CAGCTCCACCCAGATGGGCGCGGCTGCTCCATTTCCAGCTGTGACACCTCAGAG 271
QY 241 CCGTGTGGCCCAAGCATGACAGACGATGATAAAATTTCCAGTACTTGGAGAAAGT 300
Db 272 CCGTGTGGCCCAAGCATGACAGACGATGATAAAATTTCCAGTACTTGGAGAAAGT 331
QY 301 GAAAGTCCAGGGGTTTAAAAATGGGCCACTTCTCTCCAGTCCCTCTGAGGCTCCG 360
Db 332 GAAAGTCCAGGGGTTTAAAAATGGGCCACTTCTCTCCAGTCCCTCTGAGGCTCTCG 391
QY 361 CTCTGGGCGCTGCCATCTCTGCTGCTGGGCTCGGCTGCTGCTGCTGCTGCTCAT 420
Db 392 CTCTGGGCGCTGCCATCTCTGCTGCTGGGCTCGGCTGCTGCTGCTGCTGCTCAT 451
QY 421 CTGTGTGGTTGGATTTCCAAAATTTCCAGAGGACCTGCTGCTGCTGCTGCTGCTGCT 480
Db 452 CTGTGTGGTTGGATTTCCAAAATTTCCAGAGGACCTGCTGCTGCTGCTGCTGCTGCT 511
QY 481 TTTTAGCAACTTCACTCAAACTGTGTGGCGAGATCCAGGCACTGCTTCCAGGGCAG 540
Db 512 TTTTAGCAACTTCACTCAAACTGTGTGGCGAGATCCAGGCA-TCATCTCCAGGGCAG 570
QY 541 CAGCTTGGAGAAACGATAGCATCTCTGAAAGCTGAGGTGGAGGGTTTCAACGAGGAACG 600
Db 571 CAGCTTGGAGAAACGATAGCATCTCTGAAAGCTGAGGTGGAGGGTTTCAACGAGGAACG 630
QY 601 GCAGGCGAGGGGTATCTGAGCTCCAGGAACACACTACGACAGAGGACACCTAGGCCACTG 660
Db 631 GCAGG----- 635
QY 661 TCCCCACTGCCCATCTGTGTGTGTGTCAGGTTCAATTTGAAATGCTCTCGGAGTCCAGCA 720
|||||

```

Db      636  -----CAGTTTCATTCTGAAATGCTCTCGAGTCCAGCA 669
Qy      721  GCTGTCACAGACTGAAGAACTGACCTGCAGTGGCTACTCTTCAACAACTAT-CCCT 779
Db      670  GCTGTCACAGACTGAAGAACTGACCTGCAGTGGCTACTCTTCAACAACTATGCGCT 729
Qy      780  CCACTGAAGGACCTGCTGCCCGCTCACTGAGTGGGAGCACCAGACAGCTGCTACTGGT 839
Db      730  CCACTGAAGGACCTGCTGCCCGCTCACTGAGTGGGAGCACCAGACAGCTGCTACTGGT 789
Qy      840  TCTCTCACTTGGGATCTCTGGCCCGAGCTGAGAACTACTGCCAGCTGAAGAACGCC 899
Db      790  TCTCTCACTTGGGATCTCTGGCCCGAGCTGAGAACTACTGCCAGCTGAAGAACGCC 849
Qy      900  ACTGCTGCTCATCACTCTCAGGAGGAGCA 930
Db      850  ACCTGCTGCTCATCACTCTCAGGAGGAGCA 880

RESULT 6
AD008342
ID      AD008342 standard; cDNA; 944 BP.
XX      AC      AD008342;
XX      DT      01-JUL-2004 (first entry)
XX      DE      Human NOVX polynucleotide #45.
XX      KW      Human; NOVX; gene; ss; cardiomyopathy; atherosclerosis; hypertension;
KW      scleroderma; obesity; cancer; diabetes; haemophilia;
KW      graft-versus-host disease; AIDS; asthma; Crohn's disease;
KW      multiple sclerosis; infection; anorexia; cancer-associated cachexia;
KW      neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
KW      haematopoietic disorder; wasting disorder.
XX      OS      Homo sapiens.
XX      PN      US2004018594-A1.
XX      PD      29-JAN-2004.
XX      PF      01-MAY-2002; 2002US-00138588.
XX      PR      03-MAY-2001; 2001US-0288395P.
PR      04-MAY-2001; 2001US-0288900P.
PR      07-MAY-2001; 2001US-0289087P.
PR      14-MAY-2001; 2001US-0290753P.
PR      15-MAY-2001; 2001US-0291189P.
PR      16-MAY-2001; 2001US-0291243P.
PR      18-MAY-2001; 2001US-0292001P.
PR      21-MAY-2001; 2001US-0292374P.
PR      22-MAY-2001; 2001US-0292587P.
PR      23-MAY-2001; 2001US-0293107P.
PR      29-MAY-2001; 2001US-0294110P.
PR      30-MAY-2001; 2001US-0294434P.
PR      31-MAY-2001; 2001US-0294827P.
PR      31-JUL-2001; 2001US-0308901P.
PR      17-AUG-2001; 2001US-031388P.
PR      21-AUG-2001; 2001US-0313851P.
PR      21-AUG-2001; 2001US-0313937P.
PR      17-SEP-2001; 2001US-0322701P.
PR      17-SEP-2001; 2001US-0322802P.
PR      25-SEP-2001; 2001US-0324757P.
PR      27-SEP-2001; 2001US-0325314P.
PR      27-SEP-2001; 2001US-0325682P.
PR      21-NOV-2001; 2001US-0332129P.
PR      03-DEC-2001; 2001US-0336882P.
PR      14-DEC-2001; 2001US-0340305P.
XX      (ALSO/) ALSOBROOK J P.
PA      (ANDE/) ANDERSON D W.
PA      (BOLD/) BOLDOG F L.

```

```

PA      (BURG/) BURGESS C E.
PA      (CASM/) CASMAN S J.
PA      (CHAP/) CHAPOVAL A.
PA      (EDIN/) EDINGER S R.
PA      (GERL/) GERLACH V.
PA      (GORM/) GORMAN L.
PA      (GUNT/) GUNTHER E.
PA      (GUOX/) GUO X S.
PA      (KEKU/) KERUDA R.
PA      (LEPL/) LEPLEY D M.
PA      (LILL/) LI L.
PA      (LIUX/) LIU X.
PA      (MALV/) MALYANKAR U M.
PA      (MILL/) MILLER C E.
PA      (MILL/) MILLET I.
PA      (PADI/) PADIGARU M.
PA      (PATT/) PATTURAJAN M.
PA      (PENA/) PENA C E A.
PA      (RIEG/) RIEGER D K.
PA      (SHEN/) SHENOY S G.
PA      (SHIM/) SHIMKETS R A.
PA      (SPYT/) SPYTEK K A.
PA      (TAUP/) TAUPIER R J.
PA      (VERN/) VERNET C A M.
PA      (VOSS/) VOSS E Z.
PA      (ZERH/) ZERHUSEN B D.
XX
PI      Alsobrook JP, Anderson DW, Boldog FL, Burgess CE, Casman SJ;
PI      Chapoval A, Edinger SR, Gerlach V, Gorman L, Gunther E, Guo XS;
PI      Kekuda R, Lepley DM, Li L, Liu X, Malyankar UM, Miller CE;
PI      Millet I, Padigar M, Patturajan M, Pena CEA, Rieger DK, Shenoy SG;
PI      Shimkets RA, Spytek KA, Taupier RJ, Vernet CM, Voss EZ;
PI      Zerhusen BD;
XX      WPI; 2004-122037/12.
DR      P-PSDB; AD008343.
XX
PT      New NOVX polypeptides and nucleic acids, useful for preventing or
PT      treating NOVX-associated disorders, e.g. cancer, diabetes,
PT      atherosclerosis, asthma or AIDS, and in chromosome mapping, tissue typing
PT      or pharmacogenomics.
XX      Claim 20; SEQ ID NO 89; 219pp; English.
XX
CC      The invention relates to human NOVX polypeptides and the polynucleotides
CC      encoding them. The polypeptides, polynucleotides and antibodies that bind
CC      immunospecifically to the polypeptides are useful in diagnosing, treating
CC      or preventing NOVX-associated disorders such as cardiomyopathy,
CC      atherosclerosis, hypertension, scleroderma, obesity, cancer, diabetes,
CC      haemophilia, graft-versus-host disease, AIDS, asthma, Crohn's disease,
CC      multiple sclerosis, infections, anorexia, cancer-associated cachexia,
CC      neurodegenerative disorders (e.g. Alzheimer's disease or Parkinson's
CC      disease), haematopoietic disorders and wasting disorders. The
CC      polynucleotides are also used as hybridisation probes, in chromosome
CC      mapping and in tissue typing. The polypeptides are also useful as
CC      vaccines. This sequence represents a human NOVX polynucleotide of the
CC      invention.
XX
SQ      Sequence 944 BP; 209 A; 265 C; 286 G; 184 T; 0 U; 0 Other;
Query Match      49.2%; Score 717.6; DB 12; Length 944;
Best Local Similarity      87.2%; Pred. No. 1.8e-175;
Matches 933; Conservative      0; Mismatches      4; Indels 133; Gaps      8;
Qy      213  CTCCTTTTCAGCTGTGACACCTCAGAGCGCTGTGSCCAAGCATGACAGACGATG 272
Db      1      CTCCTTTTCAGCTGTGACACCTCAGAGCGCTGTGSCCTAAGCATGACAGACGATG 60
Qy      273  AAAAATTCAGTACTTGGAGATAAGTGAAGTCCAGGGGTTTAAAAATGGGCCACTTC 332
Db      61  AAAATTCAGTACTTGGAGATAAGTGAAGTCCAGGGGTTTAAAAATGGGCCACTTC 120
Qy      333  CTCTCAGTCCCTCTTCGACGCTCTCGGCTCGGCGCTGCCATCTCCTGCTCCCTGG 392

```


XX WPI; 2001-041426/05.
XX New polynucleotides encoding breast tumor specific proteins, useful for
PT prevention, treatment and diagnosis of breast cancer.
XX
XX Claim 25; Page 133-134; 165pp; English.
XX
XX The present invention provides the coding sequences for a number of
CC breast cancer related proteins. These can be used in vaccinations
CC against, diagnosis of and treatment of cancer, particularly breast cancer
XX
SQ Sequence 661 BP; 156 A; 199 C; 166 G; 139 T; 0 U; 1 Other;

Query Match 41.6%; Score 606.6; DB 4; Length 661;
Best Local Similarity 98.2%; Pred. No. 9.6e-147;
Matches 612; Conservative 1; Mismatches 10; Indels 0; Gaps 0;

QY 1 GTTGAGGAGATGGATGCTCCAGATGATAGGGCTCTGGGATTTTCAGACCCCAAGCCAGC 60
DB |||||||
QY 61 AGGACTCCAGTCACTCTACCCAGCTCTCCAGGACACAGCGCTCCCAACTCTGAGTGAC 120
DB |||||||
QY 121 GTCCCACTCTGCTCTTGGAGACAAACCAACGTTGGGAATCAACCTCCAGACCTCCCA 180
DB |||||||
QY 181 CAGCTCCACCCAGATGGGGCGGCGCTCCCTCCATTTTCAGCTGTGACAACTCAGAG 240
DB |||||||
QY 241 CCCTGTTGGCCCAAGCATCAAGGACGTATGAAAACCTTCAGTACTTGGAGATAAGGT 300
DB |||||||
QY 301 GAAAGTCAGGGGTTTAAAAATGGCCACATTCCTCTCCAGTCCCTCTGAGCGTCTCCG 360
DB |||||||
QY 361 CTCTGGCCCTGTCATCTCTGCTGCTCCCTGGGCTGGCTGCTGCTGCTGCTCATCAT 420
DB |||||||
QY 421 CTGTGCTGTTGGATTCCAAAATTCAAAATTTTCAGAGGACCTGTGTGACCTTGAGAACAG 480
DB |||||||
QY 481 TTTTAGCAACTTCACTCAAAACATGTGGCGGAGATCCAGGCACTGTACTTCCAGGGCAG 540
DB |||||||
QY 541 CAGCTTGGAGAAACGATAGCATCTCTGAAGCTGAGTGGAGGTTTCAAGCAGGAAACG 600
DB |||||||
QY 601 GCAGGCGAGGGTATCTCAGCTCC 623
DB |||||||
QY 601 GCAGGCGAGGGTATCTCAGAAATGC 623

RESULT 10
ID ABT42075 standard; DNA; 1358 BP.
XX
AC ABT42075;
XX
DT 26-JUN-2003 (first entry)
XX
DE Toxicity modelling related rat gene SEQ ID No 1777.
XX

KW Toxic effect; gene expression profile; renal toxicity; toxicity marker;
XX database; drug screening; toxicity assay; rat; db.
OS Rattus norvegicus.
XX WO200295000-A2.
XX
XX 28-NOV-2002.
XX
XX 22-MAY-2002; 2002WO-US016173.
XX
XX 22-MAY-2001; 2001US-0292335P.
XX 13-JUN-2001; 2001US-0297523P.
XX 13-JUN-2001; 2001US-0298252P.
XX 10-JUL-2001; 2001US-0303807P.
XX 10-JUL-2001; 2001US-0303808P.
XX 10-JUL-2001; 2001US-0303810P.
XX 28-AUG-2001; 2001US-0315047P.
XX 27-SEP-2001; 2001US-0324928P.
XX 22-OCT-2001; 2001US-0330462P.
XX 01-NOV-2001; 2001US-0330867P.
XX 21-NOV-2001; 2001US-0331805P.
XX 06-DEC-2001; 2001US-0336144P.
XX 19-DEC-2001; 2001US-0340873P.
XX 21-FEB-2002; 2002US-0357842P.
XX 21-FEB-2002; 2002US-0357843P.
XX 21-FEB-2002; 2002US-0357844P.
XX 15-MAR-2002; 2002US-0364134P.
XX 08-APR-2002; 2002US-0370144P.
XX 08-APR-2002; 2002US-0370206P.
XX 08-APR-2002; 2002US-0370247P.
XX 17-APR-2002; 2002US-0372794P.
XX 21-APR-2002; 2002US-0371679P.
XX
(GENE-) GENE LOGIC INC.
XX
XX Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Elashoff M;
XX WPI; 2003-148464/14.
XX
XX Predicting at least one toxic effect of a compound, useful for toxicity
PT modeling, comprises preparing a gene expression profile of a tissue or
PT cell sample exposed to the compound, and comparing the gene expression
PT profile to a database.
XX
XX Example 4; Page; 446pp; English.
XX
XX The invention relates to a novel method of predicting at least one toxic
CC effect of a compound. The method comprises a gene expression profile of a
CC tissue or cell sample exposed to the compound, and comparing the gene
CC expression profile to a database comprising at least part of the data or
CC information given in the specification. The methods are useful for
CC predicting at least one toxic effect of a compound, predicting the
CC progression of a toxic effect of a compound, predicting the renal
CC toxicity of a compound, or identifying toxicity markers in tissues or
CC cells exposed to known renal toxin. The genes are useful as toxicity
CC markers in drug screening and toxicity assays, in monitoring disease or
CC physiological states, or disease progression. This polynucleotide
CC represents a rat DNA sequence relating to the toxic effect database
CC described in the specification. NOTE: The sequence data for this patent
CC did not form part of the printed specification, but was obtained in
CC electronic format directly from the World Intellectual Property
CC Organization
XX
SQ Sequence 1358 BP; 340 A; 325 C; 365 G; 328 T; 0 U; 0 Other;

Query Match 30.0%; Score 438; DB 10; Length 1358;
Best Local Similarity 68.1%; Pred. No. 7.4e-103;
Matches 676; Conservative 0; Mismatches 295; Indels 21; Gaps 4;

QY 210 TGCTCCATTTTCAGCTGTGCAACCTCAGAGCC--GTGTTGGCCCAAGCATGACAAGGA 266
DB |||||||
DB 122 TGCTCGTTTCAGTTTCAGATAGCCCTAGAGGCTTGTTGTTGTCAGCATGACATGG 181

Db 581 TGGGAAGAAGATGAAGTCGCTAGAGTCCAGCTGGAGAAACAGCAGAGAGGACCTGAGTG 640
Qy 606 CAGGGGTATCTGAGCTCCAGGAACACACTACGAGAGGACACACCTAGGCCACTGTCGCC 665
Db 641 AAGATCACTCCAGCCTGCTGCTCCAGCTGAAGCAGTTCGTGTGACCTGCGGAGCCTGA 700
Qy 666 ACTGCCCATCTGTGTGTGCCAGTTTCATTCTGAATGCTCTCGAGTCCAGCAGCTGG 725
Db 701 GCTGTGATGCGGCGCTCCAGGCAATGGTAAGGAGGCGAGCCGCGCCCTCTCTGC 760
Qy 726 TGCAGACCTGAAGAACTGACTGCTGCCAGGTGCTACTCTCAACAACTATGCTCCACTG 785
Db 761 CTCGCCCTTCTGTGGCAGCGCTTAGCCCTGCGCCCGTTTCTCCCTCAGGCTCAG 820
Qy 786 AAGGAGCTGCTGCCCGCTCACTGGGTGAGCAGCAGCAAGACAGCTGCTACTGTTCTCTC 845
Db 821 AAGGACCTGCTGCCCGCTCACTGGGTGAGCAGCAGCGAGCTGCTACTGTTCTCTC 880
Qy 846 ACTCTGGGATGCTCTGGCGCGAGCTGAGAAGTACTGCCAGCTGAAGAACGCCACCTGG 905
Db 881 GCTCCGGGAAGCCTGGCTGAGCCGACACTACTCTGCCGCTGGAGGAGCGCACCTGG 940
Qy 906 TGTCTCACTCCAGGGAGGAGCAGAAATTTGTTCAGAAATATCTAGGCTCCGCATACA 965
Db 941 TGTGTGTCAGCTCCTGGGAGGAGCAGAAATTTGTTCAGCAGCAGCAGATAGGCCCTGTGAACA 1000
Qy 966 CTTGGATGGCCTCAGTGAAGCTGAGGAGCCTGGAAGTGGGTGAGTGAACAGACTATG 1025
Db 1001 CTTGGATGGCCTTCCAGCAAAACGGGCCCTTGAAGTGGGTGAGCGGAGCTAG 1060
Qy 1026 CGACCGGCTTCCAGAACTGGAAGCCAGGCGCAGCAGCAGCTGGCAGGGGCGCGGCTGG 1085
Db 1061 AGACGGGCTTCAAGAACTGGAGCCCGAGAGCGCCGACACTGTTAGCGCCAGGGCTCG 1120
Qy 1086 GTGGAGCGAGGACTGTGTCACTTCCATCCAGACGCGCAGGTGGAATGACGCTGTGCC 1145
Db 1121 GAGGAGCGAGGACTGTGCCCACTTCAACGACGACGCGCGCTGGAACGACGAGCTGTGCC 1180
Qy 1146 AGAGGCGCTTACCACTGGGTCTGGAGCTGGGCTGGGTGAGACGACCGCAGGAG 1198
Db 1181 AGAGGCGCTTACCGCTGGGTCTGGAGCAGAGCTGGACAAAGGCGCAGCCAGGAG 1233

RESULT 13

ABN95732
ID ABN95732 standard; DNA; 1277 BP.
AC ABN95732;
XX
DT 13-AUG-2002 (first entry)
DE Gene #2230 used to diagnose liver cancer.
XX
XX Gene; liver cancer; db; hepatocellular carcinoma; hepatotropic;
KW metastatic liver tumor; cytostatic; expression profile; disease state;
KW disease progression; drug toxicity; drug efficacy; drug metabolism.
XX
OS Homo sapiens.
XX
XX W020229103-A2.
XX
XX 11-APR-2002.
XX
PF 02-OCT-2001; 2001WO-05030589.
XX
PR 02-OCT-2000; 2000US-0237054P.
XX
PA (GENE-) GENE LOGIC INC.
XX
XX Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
PI WPI; 2002-426119/45.
XX
DR

XX
PT Diagnosing and detecting the progression of liver cancer, hepatocellular
PT carcinoma or metastatic liver tumor in a patient, involves detecting the
PT level of expression of two or more genes in a liver tissue sample.
XX
PS Claim 1; SEQ ID NO 2230; 298pp; English.
XX
CC The invention relates to a novel method for diagnosing and detecting the
CC progression of liver cancer, hepatocellular carcinoma or metastatic liver
CC tumor in a patient, and differentiating metastatic liver cancer from
CC hepatocellular carcinoma in a patient, involving detecting the level of
CC expression of two or more genes represented in ABN93503-ABN97455 in a
CC tissue sample. The method of the invention has hepatotropic, and
CC cytostatic activity. The method is useful for diagnosing and detecting
CC the progression of liver cancer, hepatocellular carcinoma and metastatic
CC liver carcinoma in a patient. The method is useful for identifying
CC expression profiles which serve as useful diagnostic markers as well as
CC markers that can be used to monitor disease states, disease progression,
CC drug toxicity, drug efficacy and drug metabolism. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX

Qy Sequence 1277 BP; 271 A; 394 C; 378 G; 234 T; 0 U; 0 Other;
Query Match 26.2%; Score 382; DB 6; Length 1277;
Best Local Similarity 66.1%; Pred. No. 2.4e-88;
Matches 546; Conservative 0; Mismatches 245; Indels 87; Gaps 3;
Db

Qy 224 CTGTGACAACTCAGAGCCGTGTTGGCCCAAGCAGTGAACAAGAGCGTATGAACATTCACAG 283
Db 140 CTGAGCAATCCAGGTCAGCGCCAGCCCTATCATGACCAAGAGGATATCAAGACCTTCAG 199
Qy 284 TACTGGAGATAGGTGAAGTCCAG---GGGTTTAAATATGGCCACTTCTCTCCAG 340
Db 200 CATCTGGCAATGAGGAGAGTGACCAACATCAGCTCAGAAAGGCGCACCTCTCTCCAG 259
Qy 341 TCCCTCTCTGACGCTCTCCGCTCTGGGCGCTGCGCATCTCTGCTGTCTCCCTGGGCTCCGC 400
Db 260 CCCCTCTGACGCTCTCTGCTCGGACCTCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 319
Qy 401 CTGCTGCTGCTGCTCATCATCTGTGTGTTGGATTCGAAATTCGAAATTCAGAGGAC 460
Db 320 CTCCTGCTGCTGTTGTTGTTGTTGATCGGATCCCAAACTCCACAGCTGCGAGGAG 379
Qy 461 CTGCTGACCTCTGAGAACAGATTTTAGCAACTTCACCTCAAACTGTGCGGAGATCCAG 520
Db 380 CTGCGGGGCTGAGAGAGACGTTTACGCAACTTCACAGCGAGCAGGAGCCAGGCTCAG 439
Qy 521 GCACCTGACTTCCAGGGCAGCAGCTTGAAGAAACGATAGCATCTCTGAAAGCTGAGGTG 580
Db 440 GGCTTGAGCACCAGGAGGCAATCTGGGAAGAAAGATGAGTCTGAGTCTCCAGCTG 499
Qy 581 GAGGTTTTCAAGCAGAGAACCGCAGGAGGATCTGAGCTCCAGGAAACACTACGAG 640
Db 500 GAGAAACAGCAGAGAGGACCTGAGTG----- 524
Qy 641 AAGGCACACCTAGGCCACTGTCCCACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 700
Db 525 -----AAGATCACTCCAGC 538
Qy 701 ATGCTCTCGAGTCCAGCAGCTGGTGAAGAACCTGAAGAACCTGACCTGCGAGCTGCT 760
Db 539 CTGCTGCTCCAGTGAAGCAGTTCGTGTCTGACCTGCGGAGGCTGAGCTGTGAGATGGG 598
Qy 761 ACTCTCAACAAATGCTCTCACTGAAGGAGCCTGCTGCCCGCTCAACTGGGTGAGAGC 820
Db 599 GCGCTCCAGGGCAATGGCT---CAGAAAGAGCCTGCTGCCCGCTCAACTGGGTGAGAGC 655
Qy 821 CAAGACAGCTGCTACTGTTCTCTCAGCTCTGGAGTGTCTGGGCGGAGGCTGAGAGTAC 880
Db 656 GAGCGCAGCTGCTACTGTTCTCTGCTCCGGGAAGGCTGCGCTGACGCCGACACTAC 715

Qy	881	TGCCAGCTGAAGAACGCCCACTGGTGGTCAATCAATCCAGGGAGGAGCAAAATTTGTC	940
Db	716	TGCCGGCTTGGAGACCGCGCACTGGTGGTCACTGGTGGGAGGAGCAAAATTTGTC	775
Qy	941	CAGAAATATCTTAGGCTCCGCATACACTGGATGGGCTTCAGTCAACCTGAAGGAGCCTGG	1000
Db	776	CAGCACACATAGGCCCTGTGMAACACTGGATGGGCTTCCACGACCAAAACGGGCCCTGG	835
Qy	1001	AAGTGGGTGGATGGAAACAGACTATGCCACGGCTTCCAGAACTTGGAGCCAGGCCAGCCA	1060
Db	836	AAGTGGGTGGACGGGACGGACTACGACCGGGCTTCAAGAACTTGGAGCCGGAGCAGCCG	895
Qy	1061	GACGACTGGCAGGGGCGACGGGCTGGGTGGAGGGCAGGAGCTGTGTCTCATCTCCATCCAGAC	1120
Db	896	GACGACTGGTACGGCCACGGGCTCGGAGGAGGCAGGACTGTGCCCACTTCCAGCAGCAC	955
Qy	1121	GGCAGGTGGATGACGACGCTTCGCACAGGCCCTACCACTGGGTCTGGAGGGCTGGCCCTG	1180
Db	956	GGCCGCTTGGAAACGACGACGCTCTGCCAGAGGCCCTACCCGCTGGGTCTCGGAGACAGAGCTG	1015
Qy	1181	GGTCAGACCAAGCCAGGAG	1198
Db	1016	GACAAGGCCACGACGAG	1033

RESULT 14

NC00114	
ACAS56588	
ID	ACA56588 standard; cDNA; 1277 BP.
XX	
XX	ACA56588;
XX	
DT	06-JUN-2003 (first entry)
XX	
DE	Human signalling pathway polynucleotide probe SEQ ID NO 1186.
XX	
XX	Human; probe; ss; array element; Parkinson's disease;
KW	signalling pathway population; cancer; adenocarcinoma; leukaemia;
KW	immunopathy; AIDS; asthma; neuropathy; Alzheimer's disease; microarray.
XX	
OS	Homo sapiens.
XX	
PN	US6500938-B1.
XX	
PD	31-DEC-2002.
XX	
PF	30-JAN-1998; 98US-00016434.
XX	
PR	30-JAN-1998; 98US-00016434.
XX	
PA	(INCY-) INCYTE GENOMICS INC.
XX	
FI	Au-Young J, Seilhamer JJ;
XX	
DR	WPI; 2003-352189/33.
XX	
PT	Combination of polynucleotide probes, useful as array elements in a
PT	microarray for monitoring the expression of a number of target
FT	polynucleotides.
XX	
PS	Claim 1; SEQ ID NO 1186; 65pp; English.

The invention relates to a combination which, comprises a number of polynucleotide probes comprising a sequence selected from one of the 1490 sequences mentioned in the specification. The combination is useful as an array element in a microarray for monitoring the expression of a number of target polynucleotides. The microarray is particularly useful in the diagnosis and treatment of cancer and immunopathology and neuropathology. The microarray is useful in diagnostics and treatment regimens, drug discovery and development, toxicological and carcinogenicity studies, forensics and pharmacogenomics. The microarray is also useful for monitoring progression of diseases and for developing sophisticated profiles for the effects of currently available therapeutic drugs. The combination is also useful for purifying a subpopulation of mRNAs. CDNA

and genomic fragments and in research and diagnostic applications. The array can detect changes in expression in a large number of genes coding for different signaling pathway populations which can be used to diagnose various diseases including cancer e.g. adenocarcinoma and leukemia, immunopathies e.g. AIDS and asthma, neuropathies e.g. Alzheimer's disease and Parkinson's disease. The present sequence represents a polynucleotide probe of the invention. Note: The sequence data for this patent did not form part of the printed specification but was obtained in electronic format directly from USPTO at seqdata.uspto.gov/sequence.html?DocID=06500938B1

SQ	Sequence 1277 BP; 271 A; 394 C; 378 G; 234 T; 0 U; 0 Other;
Query Match	26.2%; Score 382; DB 10; Length 1277;
Best Local Similarity	66.1%; Pred. No. 2.4e-88;
Matches 646; Conservative	0; Mismatches 245; Indels 87; Gaps 3;

Db 956 GCGCGCTGGAAACGACGACGCTCTGCCAGAGGCCCTACCGCTGGGTCTGGGAGACAGACTG 1015
Qy 1181 GGTGAGACCGCCAGGAG 1198
Db 1016 GACAGGCCCGCCAGGAG 1033

Search completed: May 27, 2005, 20:40:05
Job time : 848 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 27, 2005, 22:27:31 ; Search time 922 Seconds
(without alignments)
9708.761 Million cell updates/sec

Title: US-10-829-107-3
Perfect score: 1458
Sequence: 1 gttgaggagatgggatgtcc.....taataaaatgactcattt 1458

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5700845 seqs, 3069779757 residues

Total number of hits satisfying chosen parameters: 11401690

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/2/pubpna/PCTUS_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
- 13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
- 14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
- 15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
- 16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq.*
- 17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq.*
- 18: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq.*
- 19: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
- 20: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq.*
- 21: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
- 22: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1458	100.0	1458	9	US-09-862-802-3	Sequence 3, Appli
1458	100.0	1458	18	US-10-829-107-3	Sequence 3, Appli
1434.2	98.4	1617	9	US-09-728-952-45	Sequence 45, Appl
1218	83.5	1344	10	US-09-960-706-492	Sequence 492, App
1108	76.0	1370	9	US-09-862-802-9	Sequence 9, Appli
1108	76.0	1370	18	US-10-829-107-9	Sequence 9, Appli
1039.2	71.3	1072	17	US-10-138-588-89	Sequence 89, Appl
724.2	49.7	929	9	US-09-925-302-18	Sequence 18, Appl
724.2	49.7	929	10	US-09-925-302-18	Sequence 18, Appl
714.6	49.0	820	17	US-10-138-588-87	Sequence 87, Appl
438	30.0	1358	17	US-10-152-319A-1777	Sequence 1777, Ap

12	392	26.2	1277	9	US-09-880-107-2230	Sequence 2230, Ap
13	382	26.2	1277	17	US-10-305-720-1186	Sequence 1186, Ap
14	321.4	22.0	1055	17	US-10-236-392-9	Sequence 9, Appli
15	321.4	22.0	1300	17	US-09-880-107-3731	Sequence 3731, Ap
16	303	20.8	1112	17	US-10-236-392-7	Sequence 7, Appli
17	282.8	19.4	1309	9	US-09-880-107-2235	Sequence 2235, Ap
18	282.8	19.4	1309	17	US-10-236-392-5	Sequence 5, Appli
19	282.8	19.4	1309	17	US-10-283-975A-169	Sequence 169, App
20	246	16.9	443	9	US-09-960-352-7083	Sequence 7083, Ap
21	231.8	15.9	466	10	US-09-918-995-8784	Sequence 8784, Ap
22	223.6	15.3	1430	17	US-10-062-674-1931	Sequence 1931, Ap
23	212.4	14.6	1290	9	US-09-917-800A-1642	Sequence 1642, Ap
24	183	12.6	400	17	US-10-170-097-46	Sequence 46, Appl
25	183	12.6	400	19	US-10-926-684-46	Sequence 46, Appl
26	155.6	10.7	448	17	US-10-170-097-49	Sequence 49, Appl
27	155.6	10.7	448	19	US-10-926-684-49	Sequence 49, Appl
28	154.4	10.6	448	17	US-10-170-097-51	Sequence 51, Appl
29	154.4	10.6	448	19	US-10-170-097-51	Sequence 51, Appl
30	154.4	10.6	448	17	US-10-926-684-51	Sequence 51, Appl
31	154.4	10.6	448	19	US-10-926-684-51	Sequence 51, Appl
32	154	10.6	448	17	US-10-170-097-50	Sequence 50, Appl
33	154	10.6	448	19	US-10-926-684-50	Sequence 50, Appl
34	146	10.0	402	9	US-09-960-352-9823	Sequence 9823, Ap
35	142.4	9.8	410	9	US-09-960-352-11324	Sequence 11324, A
36	142	9.7	463	10	US-09-918-995-785	Sequence 785, App
37	126.2	8.7	220	17	US-10-062-674-1353	Sequence 1353, Ap
38	118	8.1	761	13	US-10-027-632-17618	Sequence 17618, A
39	118	8.1	761	13	US-10-027-632-151016	Sequence 151016, A
40	118	8.1	761	17	US-10-027-632-17618	Sequence 17618, A
41	118	8.1	761	17	US-10-027-632-151016	Sequence 151016, A
42	116.8	8.0	421	10	US-09-918-995-8213	Sequence 8213, Ap
43	115.4	7.9	402	9	US-09-960-352-2730	Sequence 2730, Ap
44	114.2	7.8	482	10	US-09-918-995-30378	Sequence 30378, A
45	109	7.5	401	9	US-09-960-352-6627	Sequence 6627, Ap

ALIGNMENTS

RESULT 1

US-09-862-802-3
; Sequence 3, Application US/09862802
; Patent No. US20020165346A1
; GENERAL INFORMATION:
; APPLICANT: Schering-Plough Corporation
; TITLE OF INVENTION: MAMMALIAN MEMBRANE PROTEIN GENES; RELATED REAGENTS
; FILE REFERENCE: SP0695B
; CURRENT APPLICATION NUMBER: US/09/862,802
; CURRENT FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 09/111,470
; PRIOR FILING DATE: 1998-07-08
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: mammalian nucleic acid
; NAME/KEY: CDS
; LOCATION: 257...1204
; OTHER INFORMATION: protein coding sequence
; NAME/KEY: variation
; LOCATION: 608...673
; OTHER INFORMATION: short form lacks these nucleotides
US-09-862-802-3

Query Match 100.0%; Score 1458; DB 9; Length 1458;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1458; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTTGAGGAGTGGGATGCCAGATGATAGGGCTCTGGGATTCAGACCCAGACGAC 60
|||||

Db 1 GTTGGAGATGGGATGTCCCAAGATAGGGCTCCTGGGATTTCCAGACCCCAAGACCGC 60
QY 61 AGGACTCCAGTCACTCTACCCAGCTCTCCAGGACACAGCGCTCCCAACTCTGAGTGAC 120
Db 61 AGGACTCCAGTCACTCTACCCAGCTCTCCAGGACACAGCGCTCCCAACTCTGAGTGAC 120
QY 121 GTCCCACTCTGTCTTTCGAGCACAAACGATGGGAATCAACGCTCCAGACCTCCCA 180
Db 121 GTCCCACTCTGTCTTTCGAGCACAAACGATGGGAATCAACGCTCCAGACCTCCCA 180
QY 181 CAGCTCCACCCAGACTGGGCGCGCGCTCTCCATTTTCAGCTGTGACACCTCCAG 240
Db 181 CAGCTCCACCCAGACTGGGCGCGCGCTCTCCATTTTCAGCTGTGACACCTCCAG 240
QY 241 CCGTGTGGCCCAAGCATCAACAGGACGTATGAAAATTTCCAGTACTTGGAGAAATAGGT 300
Db 241 CCGTGTGGCCCAAGCATCAACAGGACGTATGAAAATTTCCAGTACTTGGAGAAATAGGT 300
QY 301 GAAAGTCCAGGGGTTTAAAAATGGGCCACTTCCCTCTCCAGTCCCTCCAGCGCTCCG 360
Db 301 GAAAGTCCAGGGGTTTAAAAATGGGCCACTTCCCTCTCCAGTCCCTCCAGCGCTCCG 360
QY 361 CTCCTGGGCGCTGCATCTCTGTCTGTCTCCCTGGGCGCTCGGCTGCTGCTGCTCAT 420
Db 361 CTCCTGGGCGCTGCATCTCTGTCTGTCTCCCTGGGCGCTCGGCTGCTGCTGCTCAT 420
QY 421 CTGTGTGGTTGGATTCAAAATTTCCAAAATTTCCAGAGGACCTGGTGACCTCAGAACAGA 480
Db 421 CTGTGTGGTTGGATTCAAAATTTCCAAAATTTCCAGAGGACCTGGTGACCTCAGAACAGA 480
QY 481 TTTTAGCAACTTCACCTCAACACTGTGGCGAGATCCAGGCACTGACATTCACAGGGCAG 540
Db 481 TTTTAGCAACTTCACCTCAACACTGTGGCGAGATCCAGGCACTGACATTCACAGGGCAG 540
QY 541 CAGCTTGGAGAAACATGACATCTCTGAAAGCTGAGGTGGAGGTTTCAAGCAGGAACG 600
Db 541 CAGCTTGGAGAAACATGACATCTCTGAAAGCTGAGGTGGAGGTTTCAAGCAGGAACG 600
QY 601 GCAGGCAGGGGTATCTGAGCTCCAGAACACACTAGCAGAGGACACCTAGGCCACTG 660
Db 601 GCAGGCAGGGGTATCTGAGCTCCAGAACACACTAGCAGAGGACACCTAGGCCACTG 660
QY 661 TCCCACTGCCCCTATCTGTGTGTCTCCAGTTCATTTCTGAAATGCTCTCGAGTCCAGCA 720
Db 661 TCCCACTGCCCCTATCTGTGTGTCTCCAGTTCATTTCTGAAATGCTCTCGAGTCCAGCA 720
QY 721 GCTGTGCAAGACTGGAAGAACTGACCTGCCAGGTGGCTACTCTCAACAAACAATGCCCTC 780
Db 721 GCTGTGCAAGACTGGAAGAACTGACCTGCCAGGTGGCTACTCTCTCAACAAACAATGCCCTC 780
QY 781 CACTGAAGGACCTGCTGCCCGTCAACTGGGTGGAGACCAAGACAGCTGCTACTGGTT 840
Db 781 CACTGAAGGACCTGCTGCCCGTCAACTGGGTGGAGACCAAGACAGCTGCTACTGGTT 840
QY 841 CTCTCACTCTGGGATCTCTGGGCGAGGCTGAGAAAGTACTGCCAGCTGAAGAACGCCCA 900
Db 841 CTCTCACTCTGGGATCTCTGGGCGAGGCTGAGAAAGTACTGCCAGCTGAAGAACGCCCA 900
QY 901 CTTGGTGGTCACTCAACTCAGGAGGAGCAGAAATTTTGTCCAGAAATATCTAGGCTCCGC 960
Db 901 CTTGGTGGTCACTCAACTCAGGAGGAGCAGAAATTTTGTCCAGAAATATCTAGGCTCCGC 960
QY 961 ATACACTGGATGGGCTCAGTGACCTCAAGGACCTGGAAGTGGGTGGATGGAAACAGA 1020
Db 961 ATACACTGGATGGGCTCAGTGACCTCAAGGACCTGGAAGTGGGTGGATGGAAACAGA 1020
QY 1021 CTATGCGACCGGCTTCAGAACTGGAGCCAGCCAGCAGACTGGCAGGGGACCG 1080
Db 1021 CTATGCGACCGGCTTCAGAACTGGAGCCAGCCAGCAGACTGGCAGGGGACCG 1080
QY 1081 GCTGGGTGGAGGAGGACTGTGCTCATCTTCATCCAGAGCGGAGGTGGAATGACAGCT 1140
Db 1081 GCTGGGTGGAGGAGGAGGACTGTGCTCATCTTCATCCAGAGCGGAGGTGGAATGACAGCT 1140

RESULT 2

US-10-829-107-3

; Sequence 3, Application US/10829107

; Publication No. US20040192892A1

; GENERAL INFORMATION:

; APPLICANT: Valladeau, Jenny

; APPLICANT: Ravel, Odile

; APPLICANT: Bates, Elizabeth Ester Mary

; APPLICANT: Ford, John

; APPLICANT: Lebecque, Serge J.E.

; APPLICANT: Saeland, Sem

; TITLE OF INVENTION: Isolated Mammalian Membrane Protein Genes; Related Reagents

; FILE REFERENCE: SF0695 B

; CURRENT APPLICATION NUMBER: US/10/829,107

; PRIOR FILING DATE: 2004-04-21

; PRIOR APPLICATION NUMBER: US/09/862,802A

; PRIOR FILING DATE: 2001-05-22

; PRIOR APPLICATION NUMBER: US 60/053,080

; PRIOR FILING DATE: 1997-07-09

; PRIOR APPLICATION NUMBER: US 09/111,470

; PRIOR FILING DATE: 1998-07-08

; NUMBER OF SEQ ID NOS: 13

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 3

; LENGTH: 1458

; TYPE: DNA

; ORGANISM: Unknown

; FEATURE:

; OTHER INFORMATION: mammalian nucleic acid and protein

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (257)..(1204)

; OTHER INFORMATION: protein coding sequence

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (608)..(673)

; OTHER INFORMATION: short form lacks nucleotides 608-673

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (775)..(776)

; OTHER INFORMATION: ASGPRm (Table 2) has sequence insert encoding GEE between nucleot

; OTHER INFORMATION: ides 775-776

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (1064)..(1064)

; OTHER INFORMATION: nucleotide 1064 of DCMP2s may be A, which would encode Asn rather

; OTHER INFORMATION: than Asp at the residue numbered 270

; FEATURE:

US-10-829-107-3

Query Match 100.0%; Score 1458; DB 18; Length 1458;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1458; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 GTTGAGGAGATGGGATCTCCAGATGATAGGGCTCTGGGATTTCCAGACCCAGACGAGC 60
DB 1 GTTGAGGAGATGGGATCTCCAGATGATAGGGCTCTGGGATTTCCAGACCCAGACGAGC 60

QY 61 AGGACTCCAGTCACTCTACCCAGCTCTCCAGGACACAGGGCTCCCAACTCTGAGTGAC 120
DB 61 AGGACTCCAGTCACTCTACCCAGCTCTCCAGGACACAGGGCTCCCAACTCTGAGTGAC 120

QY 121 GTCCACCTCTGTCTCTTCCAGACACAAACGAGTGGGAATCAACCTCCAGACCTCCCA 180
DB 121 GTCCACCTCTGTCTCTTCCAGACACAAACGAGTGGGAATCAACCTCCAGACCTCCCA 180

QY 181 CAGTCCACCCAGACCTGGGGCGGGCCCTCCCTCCATTTTCAGTGTGACAACTCAGAG 240
DB 181 CAGTCCACCCAGACCTGGGGCGGGCCCTCCCTCCATTTTCAGTGTGACAACTCAGAG 240

QY 241 CCGTGTGGCCCAAGCATGACAGGAGTATGAAAACCTCCAGTACTTGGAGAAATAGGT 300
DB 241 CCGTGTGGCCCAAGCATGACAGGAGTATGAAAACCTCCAGTACTTGGAGAAATAGGT 300

QY 301 GAAAGTCCAGGGGTTTAAATAATGGGCCACTTCTCTCCAGTCCCTCTCCAGGCTCTCG 360
DB 301 GAAAGTCCAGGGGTTTAAATAATGGGCCACTTCTCTCCAGTCCCTCTCCAGGCTCTCG 360

QY 361 CTCTGGGCCCTGACATCTCTGCTGTCTCCCTGGGGCTCGGGCTGCTGCTGTGTCATCAT 420
DB 361 CTCTGGGCCCTGACATCTCTGCTGTCTCCCTGGGGCTCGGGCTGCTGCTGTGTCATCAT 420

QY 421 CTGTGTGGTGGATTCCAAATTTCCAAATTTCCAGAGGACCTGTGTACCTCGAAGACAGA 480
DB 421 CTGTGTGGTGGATTCCAAATTTCCAAATTTCCAGAGGACCTGTGTACCTCGAAGACAGA 480

QY 481 TTTTAGCAATTCACCTCAAACTCTGTGGCGAGATCCAGGCACTGACTTCCAGGGCAG 540
DB 481 TTTTAGCAATTCACCTCAAACTCTGTGGCGAGATCCAGGCACTGACTTCCAGGGCAG 540

QY 541 CAGTTGGAGAAACGATAGCATCTCTGAAAGCTGAGGTGGAGGTTTCAAGCAGGAGC 600
DB 541 CAGTTGGAGAAACGATAGCATCTCTGAAAGCTGAGGTGGAGGTTTCAAGCAGGAGC 600

QY 601 GCAGGACAGGGGTATCTGAGCTCCAGGAACACTACGAGAAAGCACACCTAGGCCACTG 660
DB 601 GCAGGACAGGGGTATCTGAGCTCCAGGAACACTACGAGAAAGCACACCTAGGCCACTG 660

QY 661 TCCCCACTGCCCATCTGTGTGTGTCCAGTTCATTTCTGAAATGCTCTGCGAGTCCAGCA 720
DB 661 TCCCCACTGCCCATCTGTGTGTGTCCAGTTCATTTCTGAAATGCTCTGCGAGTCCAGCA 720

QY 721 GCTGCTCAGACCTGAGAACTGACCTGCCAGTGGCTACTCTCAACAACTATGCTCTC 780
DB 721 GCTGCTCAGACCTGAGAACTGACCTGCCAGTGGCTACTCTCAACAACTATGCTCTC 780

QY 781 CACTGAAGGACCTGTCTGCCCTCCCTGAGGAGACCAAGACAGCTGCTACTGGTT 840
DB 781 CACTGAAGGACCTGTCTGCCCTCCCTGAGGAGACCAAGACAGCTGCTACTGGTT 840

QY 841 CTCTCACTCTGGGATGCTCTGGGCGAGGCTGAGAACTACTGCGAGCTGAAGAACGCCCA 900
DB 841 CTCTCACTCTGGGATGCTCTGGGCGAGGCTGAGAACTACTGCGAGCTGAAGAACGCCCA 900

QY 901 CTTGGTGGTCACTCACTCCAGGGAGGAGCAGAAATTTTGTCCAGAAATATCTAGGCTCCG 960
DB 901 CTTGGTGGTCACTCACTCCAGGGAGGAGCAGAAATTTTGTCCAGAAATATCTAGGCTCCG 960

QY 961 ATACACCTGGATGGGCTCAGTGACCTCTGAGGAGCCTGGAAGTGGGTGGATGGAACAGA 1020
DB 961 ATACACCTGGATGGGCTCAGTGACCTCTGAGGAGCCTGGAAGTGGGTGGATGGAACAGA 1020
```

```
QY 1021 CTATGCGACCGGCTTCCAGAACTGGAAAGCCAGGCCAGACGACTGGCAGGGGACCG 1080
DB 1021 CTATGCGACCGGCTTCCAGAACTGGAAAGCCAGGCCAGACGACTGGCAGGGGACCG 1080

QY 1081 GCTGGGTGGAGGGGAGGACTGTGCTCACTTCCATCCAGACGGCAGGTGGAAATGACGACGT 1140
DB 1081 GCTGGGTGGAGGGGAGGACTGTGCTCACTTCCATCCAGACGGCAGGTGGAAATGACGACGT 1140

QY 1141 CTGCCAGAGGCCCTTACCACTGGGCTCTGCGAGGCTGGCTGGGTCCAGACCCAGGAGAG 1200
DB 1141 CTGCCAGAGGCCCTTACCACTGGGCTCTGCGAGGCTGGCTGGGTCCAGACCCAGGAGAG 1200

QY 1201 TCACTGAGCTGCCCTTTGGTGGGACCAACCGGCCACAGAAATGCGGGTGGGAGGAGACTC 1260
DB 1201 TCACTGAGCTGCCCTTTGGTGGGACCAACCGGCCACAGAAATGCGGGTGGGAGGAGACTC 1260

QY 1261 TTCTCAGACCTCTCTCGCAAGACCGCTCTGGGAGAGAAATTAAGCACTGGGAGATTGGAAAG 1320
DB 1261 TTCTCAGACCTCTCTCGCAAGACCGCTCTGGGAGAGAAATTAAGCACTGGGAGATTGGAAAG 1320

QY 1321 CACTGCTAACTTTTGAATTTTTTCTCTTTAAATTTTAAAGATGGTATAGTGTCTTA 1380
DB 1321 CACTGCTAACTTTTGAATTTTTTCTCTTTAAATTTTAAAGATGGTATAGTGTCTTA 1380

QY 1381 AGCTTTTATTTTTTCCAACTTTTGAAAGTCAACTTCATGAAGGTATATTTTACATA 1440
DB 1381 AGCTTTTATTTTTTCCAACTTTTGAAAGTCAACTTCATGAAGGTATATTTTACATA 1440

QY 1441 ATAAATGCACTCAATTT 1458
DB 1441 ATAAATGCACTCAATTT 1458
```

RESULT 3

US-09-728-952-45

; Sequence 45, Application US/09728952
; Patent No. US20020111302A1
; GENERAL INFORMATION:
; APPLICANT: Zhou, Y. Tom
; APPLICANT: Zhou, Ping
; APPLICANT: Goodrich, Ryle
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Wang, Dunrui
; APPLICANT: Yamazaki, Vicki
; APPLICANT: Ujwal, Manusha L.
; APPLICANT: Dmanac, Radoje T.
; TITLE OF INVENTION: No. US20020111302A1el Nucleic Acids and
; FILE REFERENCE: Polypeptides
; CURRENT APPLICATION NUMBER: US/09/728,952
; CURRENT FILING DATE: 2000-11-30
; NUMBER OF SEQ ID NOS: 101
; SOFTWARE: pt_fl_genes Version 2.0
; SEQ ID NO 45
; LENGTH: 1617
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (397) .. (1356)
US-09-728-952-45

Query Match 98.4%; Score 1434.2; DB 9; Length 1617;

Best Local Similarity 99.2%; Pred. No. 0;

Matches 1455; Conservative 0; Mismatches 3; Indels 9; Gaps 1;

```
QY 1 GTTGAGGAGATGGGATGTCCAGATGATAGGGCTCTGGGATTTCCAGACCCAGACGAGC 60
DB 141 GTTGAGGAGATGGGATGTCCAGATGATAGGGCTCTGGGATTTCCAGACCCAGACGAGC 200
```

QY 61 AGGACTCCAGTCACTCTACCCAGCTCTCCAGGACACAGCGCTCCCAACTCTGAGTGAC 120
DB 201 AGGACTCCAGTCACTCTACCCAGCTCTCCAGGACACAGCGCTCCCAACTCTGAGTGAC 260
QY 121 GTCCCACTCTGCTTGTGAGGACAAACAAAGTGGGAATCAACCTCCAGCTCCCA 180
DB 261 GTCCCACTCTGCTTGTGAGGACAAACAAAGTGGGAATCAACCTCCAGCTCCCA 320
QY 181 CAGCTCCACCCAGAGCTGGGCGCGCGCTCTCCATTTTCACTGTGACAACTTCAGAG 240
DB 321 CAGCTCCACCCAGAGCTGGGCGCGCGCTCTCCATTTTCACTGTGACAACTTCAGAG 380
QY 241 CGGTGTGGCCCAAGCATGACAGGAGCTGTAAGAACTTCCAGTACTTGGAGAAAGGT 300
DB 381 CGGTGTGGCCCAAGCATGACAGGAGCTGTAAGAACTTCCAGTACTTGGAGAAAGGT 440
QY 301 GAAAGTCCAGGGGTTTAAAAATGGGCACTTCTCTCAGTCTGCTCTGCGAGGCTCTCG 360
DB 441 GAAAGTCCAGGGGTTTAAAAATGGGCACTTCTCTCAGTCTGCTCTGCGAGGCTCTCG 500
QY 361 CTCTGGGCGCTGCACTCTCTGCTGTCTGCGGCTCTGCGGCTCTGCTGCTGCTCATCAT 420
DB 501 CTCTGGGCGCTGCACTCTCTGCTGTCTGCGGCTCTGCGGCTCTGCTGCTGCTCATCAT 560
QY 421 CTGTGTGGTGGATTCAAAAATTCAAAAATTTTCCAGAGGACCTGGTGACCTTCAGAACAGA 480
DB 561 CTGTGTGGTGGATTCAAAAATTCAAAAATTTTCCAGAGGACCTGGTGACCTTCAGAACAGA 620
QY 481 TTTTAGCACTTCACTTCAAACTGTGGGGGAGATCCAGGCACTGACTTCCAGGGGAG 540
DB 621 TTTTAGCACTTCACTTCAAACTGTGGGGGAGATCCAGGCACTGACTTCCAGGGGAG 680
QY 541 CAGTTTGAAGAAACCATAGCATCTCTGAAAGCTGAGGTGGAGGTTTCAAGCAGGAAGC 600
DB 681 CAGTTTGAAGAAACCATAGCATCTCTGAAAGCTGAGGTGGAGGTTTCAAGCAGGAAGC 740
QY 601 GCAGGAGGGGTATCTGAGCTCCAGGAACACACTACGAGAAAGGACACTTAGGCCACTG 660
DB 741 GCAGGAGGGGTATCTGAGCTCCAGGAACACACTACGAGAAAGGACACTTAGGCCACTG 800
QY 661 TCCCACTGCCCATCTGTGTGTGCCAGTTTCACTTGAATGCTCTCTGAGTTCAGCA 720
DB 801 TCCCACTGCCCATCTGTGTGTGCCAGTTTCACTTGAATGCTCTCTGAGTTCAGCA 860
QY 721 GCTGTGTGAAGACCTCAAGAACTGACCTGCGAGGTGGCTACTCTCAACAACAAT- 775
DB 861 GCTGTGTGAAGACCTCAAGAACTGACCTGCGAGGTGGCTACTCTCAACAAATGCTGA 920
QY 776 ----GCTTCCACTGAAGGACCTGCTGCGCCCTCAACTGGGTGGAGCAACAGACAGCTG 831
DB 921 GGAAGCCTCCACTGAAGGACCTGCTGCGCTGTCACTGGGTGGAGCAACAGACAGCTG 980
QY 832 CTACTGGTCTCTCACTCTGGGATGCTCTGGGCGAGGCTGAGAGTACTGCGAGCTGAA 891
DB 981 CTACTGGTCTCTCACTCTGGGATGCTCTGGGCGAGGCTGAGAGTACTGCGAGCTGAA 1040
QY 892 GAAAGCCCACTGGTGTGATCACTCACTCCAGGAGGAGCAATTTTGTCCAGAAATATCT 951
DB 1041 GAAAGCCCACTGGTGTGATCACTCACTCCAGGAGGAGCAATTTTGTCCAGAAATATCT 1100
QY 952 AGGCTCCGCATACACCTGGATGGGCTCTAGTGAACCTGGAAGGCTGGAAGTGGGTGA 1011
DB 1101 AGGCTCCGCATACACCTGGATGGGCTCTAGTGAACCTGGAAGGCTGGAAGTGGGTGA 1160
QY 1012 TGGAAACAGACTATGCGACCGGCTTCAGAACTGGAAGCCAGGCGCAGACGACTGGCA 1071
DB 1161 TGGAAACAGACTATGCGACCGGCTTCAGAACTGGAAGCCAGGCGCAGACGACTGGCA 1220
QY 1072 GGGGCAAGGCTGGGTGGAGGGGAGGACTGTGCTCACTTCCATTCAGAGCGGAGGTGGA 1131
DB 1221 GGGGCAAGGCTGGGTGGAGGGGAGGACTGTGCTCACTTCCATTCAGAGCGGAGGTGGA 1280
QY 1132 TGACGAGCTGTGCGAGAGGCGCTTACCACTGGGTCTGCGAGGCTGGGCTGGGCTCAGACCAG 1191

DB 1281 TGAAGACGCTCTGCCAGAGGCGCTTACCCTGAGGCTTGGGCTGAGACCCAG 1340
QY 1192 CCAGGAGAGTCACTGAGCTGCTTGTGGGACCAACCGGCGCAGAGAAATGCGGCTGGGA 1251
DB 1341 CCAGGAGAGTCACTGAGCTGCTTGTGGGACCAACCGGCGCAGAGAAATGCGGCTGGGA 1400
QY 1252 GGAGGACTCTTCTCCAGACCTCTCGAAGACCGCTCTGGGAGAGAAATAAGCACTGGGA 1311
DB 1401 GGAGGACTCTTCTCCAGACCTCTCGAAGACCGCTCTGGGAGAGAAATAAGCACTGGGA 1460
QY 1312 GATTGGAAGCACTGCTAACTTTTGAATTTTCTTTTAAATTTTAAAAAGATGTTATA 1371
DB 1461 GATTGGAAGCACTGCTAACTTTTGAATTTTCTTTTAAATTTTAAAAAGATGTTATA 1520
QY 1372 GTGTCTTTAAGCTTTTATTTTCCAACTTTTGAAGTCAACTTCAATGAGGTATAAT 1431
DB 1521 GTGTCTTTAAGCTTTTATTTTCCAACTTTTGAAGTCAACTTCAATGAGGTATAAT 1580
QY 1432 TTTTACATAATAAAAAATGCACCTCAATTT 1458
DB 1581 TTTTACATAATAAAAAATGCACCTCAATTT 1607

RESULT 4

US-09-960-706-492
; Sequence 492, Application US/09960706
; Publication No. US20030134280A1
; GENERAL INFORMATION:
; APPLICANT: Munger, William E.
; TITLE OF INVENTION: Identifying Drugs for and Diagnosis of Benign Prostatic Hyperplas
; FILE REFERENCE: 44921-5029-01US
; CURRENT APPLICATION NUMBER: US/09/960,706
; PRIOR APPLICATION NUMBER: 2001-09-24
; PRIOR FILING DATE: 60/223,323
; PRIOR APPLICATION NUMBER: 2000-08-07
; PRIOR FILING DATE: 09/873,319
; NUMBER OF SEQ ID NOS: 1124
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 492
; LENGTH: 1344
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20030134280A1 D50532
US-09-960-706-492

Query Match 83.5%; Score 1218; DB 10; Length 1344;
Best Local Similarity 93.7%; Pred. No. 0;
Matches 1328; Conservative 0; Mismatches 0; Indels 90; Gaps 2;
QY 50 CCAAGACAGCAGGACTCCAGTCACTTACCCAGCTCTCCAGGACACAGCGTCCCAA 109
DB 1 CCAAGACAGCAGGACTCCAGTCACTTACCCAGCTCTCCAGGACACAGCGTCCCAA 60
QY 110 CTCTGAGTGAGTCCCACTCTGCTTCTTGAGCACAACCAACGTTGGGAATCAACCCCTC 169
DB 61 CTCTGAGTGAGTCCCACTCTGCTTCTTGAGCACAACCAACGTTGGGAATCAACCCCTC 120
QY 170 CAGACTCCCACTGAGTCCCACTGAGTGGGCGCGGCTGCTTCCCTTCCATTTTCACTGTGA 229
DB 121 CAGACTCCCACTGAGTCCCACTGAGTGGGCGCGGCTGCTTCCCTTCCATTTTCACTGTGA 180
QY 230 CAACCTCAGAGCGGTGTTGGCCCAAGCATGACAGGAGTATGAAAACTTCAGTACTTG 289
DB 181 CAACCTCAGAGCGGTGTTGGCCCAAGCATGACAGGAGTATGAAAACTTCAGTACTTG 240
QY 290 GAGATAGGTGAAAGTCCAGGGGTTTAAAAATGGGCACCTTCTCTCCAGTCCCTCTG 349
DB 241 GAGATAGGTGAAAGTCCAGGGGTTTAAAAATGGGCACCTTCTCTCTCAGTCCCTCTG 300

Qy	628	ACACACTACG	CAGNAGCG	CACCTAG	CGCCACTGTGCCCACTGTGCTGTGTGCC	687	
Db	565	-----	-----	-----	-----C	565	
Qy	688	AGTTCATTCTG	AAATGCTCTCTG	CAGTCCAG	AGCTGGTCAAGA	CCTGAAGAAACTGAC	747
Db	566	AGTTCATTCTG	AAATGCTCTCTG	CAGTCCAG	AGCTGGTCAAGA	CCTGAAGAAACTGAC	625
Qy	748	CTGCCAGGTGG	CTACTCTCAACA	CAAT-----	GCCTCCACTGA	AGGGACCTGCTG	798
Db	626	CTGCCAGGTGG	CTACTCTCAACA	CAATGGTGA	AGGAGCCCTCCACTGA	AGGGACCTGCTG	685
Qy	799	CCCGTCAACTGG	TGGAGCACC	CAAGACGCTGCTACT	TGGTTCTCTCACTCT	CGGGATGTC	858
Db	686	CCCGTCAACTGG	TGGAGCACC	CAAGACGCTGCTACT	TGGTTCTCTCACTCT	CGGGATGTC	745
Qy	859	CTGGCCGAGGCT	GAGAACTACTGCC	AGCTGAAGA	AGCCCACTGGTGGT	CTCAACTCAATC	918
Db	746	CTGGCCGAGGCT	GAGAACTACTGCC	AGCTGAAGA	AGCCCACTGGTGGT	CTCAACTCAATC	805
Qy	919	CAGGGAGGAGCA	GAATTTTGTCTC	AGAATATCTAG	GCTCCGATACACCT	TGGAATGGGCCCT	978
Db	806	CAGGGAGGAGCA	GAATTTTGTCTC	AGAATATCTAG	GCTCCGATACACCT	TGGAATGGGCCCT	865
Qy	979	CAGTGACCTTGA	AGGAGCCTGG	AGTGGGTGA	TGGAACAGACTATG	CGACCGGCTTCCA	1038
Db	866	CAGTGACCTTGA	AGGAGCCTGG	AGTGGGTGA	TGGAACAGACTATG	CGACCGGCTTCCA	925
Qy	1039	GAACCTGGAAG	CCAGGCCAG	CAGACGACTGG	CAGGGCAGCGGCTGGGT	TGGAGCGAGGA	1098
Db	926	GAACCTGGAAG	CCAGGCCAG	CAGACGACTGG	CAGGGCAGCGGCTGGGT	TGGAGCGAGGA	985
Qy	1099	CTGTGCTCACTT	CCATCCAGAC	GGCAGGTGGA	ATGAGCAGTCTTGC	CAGAGGCCCTACCA	1158
Db	986	CTGTGCTCACTT	CCATCCAGAC	GGCAGGTGGA	ATGAGCAGTCTTGC	CAGAGGCCCTACCA	1045
Qy	1159	CTGGGCTCGG	AGGCTGGCTGG	GTCCAGAC	CAGCAGGAGAGTCACT	TGAGCTGCCCTTTGG	1218
Db	1046	CTGGGCTCGG	AGGCTGGCTGG	GTCCAGAC	CAGCAGGAGAGTCACT	TGAGCTGCCCTTTGG	1105
Qy	1219	TGGGACACCC	GGCCACAGAA	ATGGCGGTGG	GAGGAGACTCTTCTCA	CGACCTCTCTGCG	1278
Db	1106	TGGGACACCC	GGCCACAGAA	ATGGCGGTGG	GAGGAGACTCTTCTCA	CGACCTCTCTGCG	1165
Qy	1279	AAGACCCCTCT	GGGAGAGAA	TAGCACTGG	GAGATTGGAAGCACTGCT	TAAGATTTTGAA	1338
Db	1166	AAGACCCCTCT	GGGAGAGAA	TAGCACTGG	GAGATTGGAAGCACTGCT	TAAGATTTTGAA	1225
Qy	1339	TTTTTTTCTCT	TTTAAATTTAA	AGATGGTATAGT	GTCTTAAAGCTTTTAT	TTTTTTTTTTC	1398
Db	1226	TTTTTTTCTCT	TTTAAATTTAA	AGATGGTATAGT	GTCTTAAAGCTTTTAT	TTTTTTTTTTC	1285
Qy	1399	AACTTTGAAG	TCACTTCAT	GAGGTATAATTTTT	TTCATAATAATAA	ATGCACTCATTT	1458
Db	1286	AACTTTGAAG	TCACTTCAT	GAGGTATAATTTTT	TTCATAATAATAA	ATGCACTCATTT	1345

RESULT 6

```

US-10-829-107-9
; Sequence 9, Application US/10829107
; Publication NO. US20040192892A1
;
; GENERAL INFORMATION:
;
; APPLICANT: Valladeau, Jenny
;
; APPLICANT: Ravel, Odile
;
; APPLICANT: Bates, Elizabeth Ester Mary
;
; APPLICANT: Ford, John
;
; APPLICANT: Lebecque, Serge J.E.
;
; APPLICANT: Saeland, Sem
;
; TITLE OF INVENTION: Isolated Mammalian Membrane Protein Genes; Related Reagents
;
; FILE REFERENCE: SF0695 B
;
; CURRENT APPLICATION NUMBER: US/10/829,107
;
; CURRENT FILING DATE: 2004-04-21
;
; PRIOR APPLICATION NUMBER: US/09/862,802A

```

Db 626 CTGCCAGGTGGCTACTCTCAACACAAATGGTGAGGAGCCCTCCACTGAAGGACCTGCTG 685
Qy 799 CCGCGTCAACTGGGTGAGACCAAGACAGCTGCTACTGTTTCTCTCACTCTGGGATGTC 858
Db 686 CCGCGTCAACTGGGTGAGACCAAGACAGCTGCTACTGTTTCTCTCACTCTGGGATGTC 745
Qy 859 CTGGGCGAGGCTCAGAAAGTACTGCCAGCTGAAGAACGCCACCTGGTGGTGCATCAACTC 918
Db 746 CTGGGCGAGGCTCAGAAAGTACTGCCAGCTGAAGAACGCCACCTGGTGGTGCATCAACTC 805
Qy 919 CAGGAGGAGCAGAAATTTGTCCAGAAATATCTAGGCTCCGCATACACCTGGATGGGCCT 978
Db 806 CAGGAGGAGCAGAAATTTGTCCAGAAATATCTAGGCTCCGCATACACCTGGATGGGCCT 865
Qy 979 CAGTGACCTTGAAGGACCTGGAAGTGGGTGGATGGAACAGACTATCGGACCGGCTCCA 1038
Db 866 CAGTGACCTTGAAGGACCTGGAAGTGGGTGGATGGAACAGACTATCGGACCGGCTCCA 925
Qy 1039 GAACCTGAAGCAGGCGCAGACGACTGGCAGGGGCACGGGCTGGGTGGAGCGGAGGA 1098
Db 926 GAACCTGAAGCAGGCGCAGACGACTGGCAGGGGCACGGGCTGGGTGGAGCGGAGGA 985
Qy 1099 CTGTGCTCACTTCCATCCAGACGCGCAGGTGGAATGACGAGCTCTGCCAGAGGCCCTACCA 1158
Db 986 CTGTGCTCACTTCCATCCAGAGGCGCAGGTGGAATGACGAGCTCTGCCAGAGGCCCTACCA 1045
Qy 1159 CTGGGCTCGGAGCTGGGCTGGGTGAGACGAGGTGGAATGACGAGCTCTGCCAGAGGCCCTACCA 1218
Db 1046 CTGGGCTCGGAGCTGGGCTGGGTGAGACGAGGTGGAATGACGAGCTCTGCCAGAGGCCCTACCA 1105
Qy 1219 TGGGACCAACCGGCGCAGAGAAATGGCGGTGGGAGGAGCTCTTCTCAAGACCTCCCTCGC 1278
Db 1106 TGGGACCAACCGGCGCAGAGAAATGGCGGTGGGAGGAGCTCTTCTCAAGACCTCCCTCGC 1165
Qy 1279 AAGACCCCTCTGGGAGAGAAATAGCACTGGGAGATTGGAGCACTGCTAACATTTTGAA 1338
Db 1166 AAGACCCCTCTGGGAGAGAAATAGCACTGGGAGATTGGAGCACTGCTAACATTTTGAA 1225
Qy 1339 TTTTCTCTTTAAATTTTAAAGATGGTATAGTGTCTTAAGCTTTTATTTTTC 1398
Db 1226 TTTTCTCTTTAAATTTTAAAGATGGTATAGTGTCTTAAGCTTTTATTTTTC 1285
Qy 1399 AACTTTTGAAGTCAACTCATGAGGTATATTTTACATATAAAATGCATCATTT 1458
Db 1286 AACTTTTGAAGTCAACTCATGAGGTATATTTTACATATAAAATGCATCATTT 1345

RESULT 7
US-10-138-588-89
; Sequence 89, Application US/10138588
; Publication No. US20040018594A1
; GENERAL INFORMATION:
; APPLICANT: Alsbrook et al.
; TITLE OF INVENTION: NOVEL ANTIBODIES THAT BIND TO ANTIGENIC POLYPEPTIDES, NUCLEIC ACID
; TITLE OF INVENTION: ENCODING THE ANTIGENS, AND METHODS OF USE
; FILE REFERENCE: 21402-347A
; CURRENT APPLICATION NUMBER: US/10/138,588
; PRIORITY FILING DATE: 2002-05-01
; PRIORITY FILING DATE: 2001-05-03
; PRIORITY FILING DATE: 2001-05-03
; PRIORITY FILING DATE: 2001-07-31
; PRIORITY FILING DATE: 2001-07-31
; PRIORITY FILING DATE: 2001-08-17
; PRIORITY FILING DATE: 2001-08-17
; PRIORITY FILING DATE: 2001-09-25
; PRIORITY FILING DATE: 2001-09-25
; PRIORITY FILING DATE: 2001-05-04
; NUMBER OF SEQ ID NOS: 203
; SEQ ID NO 89
; LENGTH: 1072
; TYPE: DNA
; ORGANISM: Homo sapiens

; FEATURE:
; NAME/KEY: CDS
; LOCATION: (45) .. (1002)
US-10-138-588-89

Query Match 71.3%; Score 1039.2; DB 17; Length 1072;
Best Local Similarity 98.9%; Pred. No. 3.2e-278;
Matches 1060; Conservative 0; Mismatches 3; Indels 9; Gaps 1;

Qy 213 CTCATTTTCAGCTGTGACAACTCAGAGCCGCTGTGTGCCCAAGCATGACAAGGACGTATG 272
Db 1 CTCATTTTCAGCTGTGACAACTCAGAGCCGCTGTGTGCCCTAAGCATGACAAGGACGTATG 60
Qy 273 AAAACTTCCAGTACTTGGAGAAATAAGGTGAAGTCCAGGGGTTTAAAAATGGGCCACTTC 332
Db 61 AAAACTTCCAGTACTTGGAGAAATAAGGTGAAGTCCAGGGGTTTAAAAATGGGCCACTTC 120
Qy 333 CTCTCAGTCCCTCTCTGACGCTCTCCGCTCTGGGCCCTGCCATCTCTGCTGTGCTCCCTGG 392
Db 121 CTCTCAGTCCCTCTCTGACGCTCTCTGCTCTGGGCCCTGCCATCTCTGCTGTGCTCCCTGG 180
Qy 393 GCCTCGGCTCTGCTGCTGCTCATCATCTGTGTGTGGATTCCAAAAATTTCCAAATTTTC 452
Db 181 GCCTCGGCTCTGCTGCTGCTCATCATCTGTGTGTGGATTCCAAAAATTTCCAAATTTTC 240
Qy 453 AGAGGACCTGTGTGACCTGAGAAACAGATTTTAGCAACTTCACTCAAACTGTGGCGG 512
Db 241 AGAGGACCTGTGTGACCTGAGAAACAGATTTTAGCAACTTCACTCAAACTGTGGCGG 300
Qy 513 AGATCCAGGCACTGACTTCCCGGCGGAGAGCTTTGGAAGAAACGATAGCATCTCTGAAAG 572
Db 301 AGATCCAGGCACTGACTTCCCGGCGGAGAGCTTTGGAAGAAACGATAGCATCTCTGAAAG 360
Qy 573 CTGAGGTGAGGGTCTTCAAGCAGGAAACGAGCAGGAGGGGTATCTGAGCTCCAGGAAACACA 632
Db 361 CTGAGGTGAGGGTCTTCAAGCAGGAAACGAGCAGGAGGGGTATCTGAGCTCCAGGAAACACA 420
Qy 633 CTAGCAGAAAGGACACCTTAGGCACTGTCCCACTGCCCCATCTGTGTGTGCTCCAGTTC 692
Db 421 CTAGCAGAAAGGACACCTTAGGCACTGTCCCACTGCCCCATCTGTGTGTGCTCCAGTTC 480
Qy 693 ATTCTGAAGTCTCTCGGAGTCCAGCAGCTGTGTGCAAGAACCTTGAAGAAACGATAGCATCTG 752
Db 481 ATTCTGAAGTCTCTCGGAGTCCAGCAGCTGTGTGCAAGAACCTTGAAGAAACGATAGCATCTG 540
Qy 753 AGGTGGTACTCTCAACAAACAAT-----GCCTCCACTGAAGGAGACCTGTGCCCCCG 803
Db 541 AGGTGGTACTCTCAACAAACAATGGTGAGGAGGCTCCACTGAAGGAGACCTGTGCCCCCG 600
Qy 804 TCAACTGGGTGAGGACCAAGACAGCTGTCTA CTGGTTCCTCTCACTCTGGGATGTCTGGG 863
Db 601 TCAACTGGGTGAGGACCAAGACAGCTGTCTA CTGGTTCCTCTCACTCTGGGATGTCTGGG 660
Qy 864 CCAGGCTGAGAGTACTGCCAGTGAAGAACCCCACTGCTGGTGTGCTCACTCAACTCCAGGG 923
Db 661 CCAGGCTGAGAGTACTGCCAGTGAAGAACCCCACTGCTGGTGTGCTCACTCAACTCCAGGG 720
Qy 924 AGGACAGAAATTTTGTCCAGAAATATCTAGGCTCCGCATACACCTGATGGGCTCAGTG 983
Db 721 AGGACAGAAATTTTGTCCAGAAATATCTAGGCTCCGCATACACCTGATGGGCTCAGTG 780
Qy 984 ACCCTGAAGAGCCTGGAGTGGGTGGATGGAAACAGACTATGCGACCGGCTTCAGAACT 1043
Db 781 ACCCTGAAGAGCCTGGAGTGGGTGGATGGAAACAGACTATGCGACCGGCTTCAGAACT 840
Qy 1044 GGAAGCCAGCCAGACCAAGACGACTGGCAGGGGCAAGGGCTGGGTGGAGGGGAGGACTGTG 1103
Db 841 GGAAGCCAGCCAGACCAAGACGACTGGCAGGGGCAAGGGCTGGGTGGAGGGGAGGACTGTG 900
Qy 1104 CTCACTTCCATCCAGAGCGGAGGTGAATGACGAGCTCTGCCAGAGGCCCTTACCACTGGG 1163
Db 901 CTCACTTCCATCCAGAGCGGAGGTGAATGACGAGCTCTGCCAGAGGCCCTTACCACTGGG 960


```
QY 1164 TCTGCGAGGCTGCCTGGGTGAGCCAGCCAGGAGAGTCACTGAGCTGCTTTGTGGGA 1223
Db 961 TCTGCGAGGCTGCCTGGGTGAGCCAGCCAGGAGAGTCACTGAGCTGCTTTGTGGGA 1020
QY 1224 CCACCCGGCCACAGAAATGGCGTGGGAGGAGCTCTTCTCAGCACTCCT 1275
Db 1021 CCACCCGGCCACAGAAATGGCGTGGGAGGAGTCTTCTCAGCACTCCT 1072

RESULT 8
US-09-925-302-18
; Sequence 18, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 929
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (431)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (613)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (918)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (929)
; OTHER INFORMATION: n equals a,t,g, or c
;
US-09-925-302-18
Query Match 49.7%; Score 724.2; DB 9; Length 929;
Best Local Similarity 89.8%; Pred. No. 1.4e-190;
Matches 836; Conservative 6; Mismatches 6; Indels 83; Gaps 3;

QY 1 GTTGAGGAGATGGGATGTCCAGATGATAGGGCTCCTGGGATTTCAAGCCCAAGACCAGC 60
Db 32 GTTGAGGAGATGGGATGTCCAGATGATAGGGCTCCTGGGATTTCAAGCCCAAGACCAGC 91
QY 61 AGGACTCCAGTCACTCTACCCAGCTCTCCAGGACACAGCGCTCCAACTCTGAGTGC 120
Db 92 AGGACTCCAGTCACTCTACCCAGCTCTCCAGGACACAGCGCTCCAACTCTGAGTGC 151
QY 121 GTCCCACTCTGTCTCTGAGCACAACCAAGTGGGAATCACACCCTCCAGACCTCCCA 180
Db 152 GTCCCACTCTGTCTCTGAGCACAACCAAGTGGGAATCACACCCTCCAGACCTCCCA 211
QY 181 CAGCTCCACCCAGACTGGCGCGCCCTGCTCCATTTTCAGCTGTGACACCTCAGAG 240
Db 212 CAGCTCCACCCAGACTGGCGCGCCCTGCTCCATTTTCAGCTGTGACACCTCAGAG 271
QY 241 CCGTGTGGCCCAAGCATGACAGGACGTATGAAAATCTTCAGTACTTGGAATAAGGT 300
Db 272 CCGTGTGGCCCAAGCATGACAGGACGTATGAAAATCTTCAGTACTTGGAATAAGGT 331
QY 301 GAAAGTCCAGGGGTTTTAAAAATGGGCCACTTCTCTCCAGTCCCTCCTCAGCGTCTCG 360
Db 332 GAAAGTCCAGGGGTTTTAAAAATGGGCCACTTCTCTCCAGTCCCTCCTCAGCGTCTCG 391
QY 361 CTCTGGGCCCCTGCCATCTCTGCTGCTCCCTGGGCCCCTCGGCCCTGCTGCTGCTCATCAT 420
```

```
Db 392 CTCTGGGSCCTGCCATCTCTCTGCTGTCCCTGGGCTCGGNCCTCCTGCTGCTCATCAT 451
QY 421 CTGTGTGTTGGATTTCCAAAATTTCCAAAATTTCCAGAGGACCTGTGTGACCTTGAGAACAGA 480
Db 452 CTGTGTGTTGGATTTCCAAAATTTCCAAAATTTCCAGAGGACCTGTGTGACCTTGAGAACAGA 511
QY 481 TTTTAGCAACTTTCACCTCAAAACACTGTGGCGGAGATCCAGGCACTGACTTCCAGGGGAG 540
Db 512 TTTTAGCAACTTTCACCTCAAAACACTGTGGCGGAGATCCAGGCA-TGACTTCCAGGGGAG 570
QY 541 CAGCTTGAAGAAACGATAGCATCTCTGAAGAGCTGAGGTGAGGGTTTCAAGACGAGAACG 600
Db 571 CAGCTTGAAGAAACGATAGCATCTCTGAAGAGCTGAGGTGAGGGTTTCAAGACGAGAACG 630
QY 601 GCAGCGAGGGGTATCTGAGCTCCAGGAACACACTACGCGAGGAAGGCACACCTAGGCCACTG 660
Db 631 GCAGS----- 635
QY 661 TCCCACACTGCCCATCTGTGTGTCTCCAGTTCAATCTGAAATGCTCCTGCGAGTCCAGCA 720
Db 636 -----CAGTTCAATCTGAAATGCTCCTGCGAGTCCAGCA 669
QY 721 GCTGTGCAAGACCTGGAAGAACTGACCTGCCAGGTGGCTACTCTCAACAAAT-GCCT 779
Db 670 GCTGTGCAAGACCTGGAAGAACTGACCTGCCAGGTGGCTACTCTCAACAAATGSGCT 729
QY 780 CCACCTGAAGGACCTGCTGCCCTCAACTGGGTGGAGCACCAAGACAGCTGCTACTTGGT 839
Db 730 CCACCTGAAGGACCTGCTGCCCTCAACTGGGTGGAGCACCAAGACAGCTGCTACTTGGT 789
QY 840 TCTCTCACTCTGGGATGCTCTGGGCGGAGGCTGAGAAAGTACTGCCAGCTGAAGAACGCC 899
Db 790 TCTCTCACTCTGGGATGCTCTGGGCGGAGGCTGAGAAAGTACTGCCAGCTGAAGAACGCC 849
QY 900 ACCTGTGTGTCATCAACTCCAGGGAGGAGCA 930
Db 850 ACCTGTGTGTCATCAAAATCCAGGGAGGAGCA 880

RESULT 9
US-09-925-302-18
; Sequence 18, Application US/09925302
; Publication No. US20030064072A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 929
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (431)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (613)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (918)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (929)
; OTHER INFORMATION: n equals a,t,g, or c
```


US-09-925-302-18

Query Match 49.7%; Score 724.2; DB 10; Length 929;
Best Local Similarity 89.8%; Pred. No. 1.4e-190;
Matches 836; Conservative 6; Mismatches 6; Indels 83; Gaps 3;

```
QY 1 GTTGAGGAGATGGGATGCTCCAGATGATAGGGCTCTCGGATTTCCAGACCCAGACGAGC 60
DB 32 GTTGAGGAGATGGGATGCTCCAGATGATAGGGCTCTCGGATTTCCAGACCCAGACGAGC 91
QY 61 AGGACTCCAGTCACTCTACCCAGCTCTCCAGGACACAGCGCTCCCAACTCTGAGTGAC 120
DB 92 AGGACTCCAGTCACTCTACCCAGCTCTCCAGGACACAGCGCTCCCAACTCTGAGTGAC 151
QY 121 GTCCCACTCTGTGCTCTTGCAGACAAACCAAGTGGGAATCACACCTCCAGACCTCCCA 180
DB 152 GTCCCACTCTGTGCTCTTGCAGACAAACCAAGTGGGAATCACACCTCCAGACCTCCCA 211
QY 181 CAGTCCACCCAGACTGGGGCGCGCCCTCCATTTTCAGCTGTGACAACTCAGAG 240
DB 212 CAGTCCACCCAGACTGGGGCGCGCCCTCCATTTTCAGCTGTGACAACTCAGAG 271
QY 241 CCCTGTTGGCCCAAGCATGACAGGAGTATGAAAACCTTCCAGTACTTGGAGAATAAGGT 300
DB 272 CCCTGTTGGCCCAAGCATGACAGGAGTATGAAAACCTTCCAGTACTTGGAGAATAAGGT 331
QY 301 GAAAGTCCAGGGGTTTAAATAATGGGCCACTTCTCTCCAGTCCCTCTCCAGGCTCTCCG 360
DB 332 GAAAGTCCAGGGGTTTAAATAATGGGCCACTTCTCTCCAGTCCCTCTCCAGGCTCTCCG 391
QY 361 CTCTGGGCGCTGCATCTCTGCTGCTCCCTGGGCGCTGCTGCTGCTGCTGCTCATCAT 420
DB 392 CTCTGGGCGCTGCATCTCTGCTGCTCCCTGGGCGCTGCTGCTGCTGCTGCTCATCAT 451
QY 421 CTCTGTTGGTGGATTTCCAAATTTCCAGAGGGAAGCTGCTGACCTTGAGAAACAGA 480
DB 452 CTCTGTTGGTGGATTTCCAAATTTCCAGAGGGAAGCTGCTGACCTTGAGAAACAGA 511
QY 481 TTTTAGCAATTCACCTCAAACTCTGTGGGAGATCCAGGCACTGACTTCCAGGGCAG 540
DB 512 TTTTAGCAATTCACCTCAAACTCTGTGGGAGATCCAGGCA-TGACTTCCAGGGCAG 570
QY 541 CAGCTTGGAGAAACGATAGCATCTGAAAGCTGAGGTGGAGGTTTCAAGCAGGAACG 600
DB 571 CAGCTTGGAGAAACGATAGCATCTGAAAGCTGAGGTGGAGGTTTCAAGCAGGAACG 630
QY 601 GCAGGAGGGGTATCTGAGCTCCAGGAACACACTACGACAGAAAGCACACCTAGGCCACTG 660
DB 631 GCAGG----- 635
QY 661 TCCCCACTGCCCATCTGTGTGTGTGCTCCAGTTCATTTCTGAAATGCTCTCGAGTCCAGCA 720
DB 636 -----CAGTTTCATTTCTGAAATGCTCTCGAGTCCAGCA 669
QY 721 GCTGTTGCAAGACTGAGAAACTGACCTGCCAGTGGCTACTCTCAACAACT-GCCT 779
DB 670 GCTGTTGCAAGACTGAGAAACTGACCTGCCAGTGGCTACTCTCAACAACTAGGCCT 729
QY 780 CCACTGAAGGACCTGCTGCCCGCTCAACTGGGTGGAGCACCAAGACAGCTGCTACTGGT 839
DB 730 CCACTGAAGGACCTGCTGCCCGCTCAACTGGGTGGAGCACCAAGACAGCTGCTACTGGT 789
QY 840 TCTCTCACTCTGGGATGCTCTGGGCGGAGGCTGAGAAAGTACTGCCAGCTGAAGAACGCC 899
DB 790 TCTCTCACTCTGGGATGCTCTGGGCGGAGGCTGAGAAAGTACTGCCAGCTGAAGAACGCC 849
QY 900 ACCTGGTGGTCACTCAACTCCAGGGAGGAGCA 930
DB 850 ACCTGGTGGTCACTCAATCCAGGGAGGAGCA 880
```

RESULT 10

US-10-138-588-87

; Sequence 87, Application US/10138588
; Publication No. US20040018594A1
; GENERAL INFORMATION:
; APPLICANT: Alsbrook et al.
; TITLE OF INVENTION: NOVEL ANTIBODIES THAT BIND TO ANTIGENIC POLYPEPTIDES, NUCLEIC ACID
; FILE REFERENCE: 21402-347A
; CURRENT APPLICATION NUMBER: US/10/138,588
; PRIOR FILING DATE: 2002-05-01
; PRIOR APPLICATION NUMBER: 60/288,395
; PRIOR FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: 60/308,901
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/313,388
; PRIOR FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: 60/324,757
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: 60/288,900
; PRIOR FILING DATE: 2001-05-04
; NUMBER OF SEQ ID NOS: 203
; SEQ ID NO 87
; LENGTH: 820
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3)..(804)
US-10-138-588-87

Query Match 49.0%; Score 714.6; DB 17; Length 820;

Best Local Similarity 90.0%; Pred. No. 6.1e-188;

Matches 811; Conservative 0; Mismatches 9; Indels 81; Gaps 1;

```
QY 321 ATGGGGCACTTCTCTCCAGTCCCTCTCGAGGCTCTCGAGGCTCTCGGCGCTCGGCGCTCCCATCTCC 380
DB 1 ATGGGGCACTTCTCTCCAGTCCCTCTCGAGGCTCTCGAGGCTCTCGGCGCTCGGCGCTCCCATCTCC 60
QY 381 TGTGTGCTGGGCGCTCGGCGCTCTCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 440
DB 61 TGTGTGCTGGGCGCTCGGCGCTCTCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 120
QY 441 ATTCCAAATTTCCAGAGGACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 500
DB 121 ATTCCAAATTTCCAGAGGACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 180
QY 501 ACATGTGGCGGAGATCCAGGCACTGACTTCCAGGCGCAGCAGCTTGGAGAAACGATAG 560
DB 181 ACATGTGGCGGAGATCCAGGCACTGACTTCCAGGCGCAGCAGCTTGGAGAAACGATAG 240
QY 561 CATCTGTGAAAGCTGAGGTGGAGGGTTTCAAGCAGGAAACGCGGATCTGTGAGC 620
DB 241 CATCTGTGAAAGCTGAGGTGGAGGGTTTCAAGCAGGAAACGCGGATCTGTGAGC 285
QY 621 TCCAGGACACACTACGACAGAGGACACCTAGGCCACTGTGCCCATCTGTGT 680
DB 286 ----- 285
QY 681 GTGTCCAGTTCATTTCTGAAATGCTCTCGAGTCCAGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 740
DB 286 -----CAGTTTCATTTCTGAAATGCTCTCGAGTCCAGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 339
QY 741 AACTGACCTGCCAGGTGGTGTCTCTCAACAAATGCTTCCAGTGAAGGAGCTGTGTGTGTGTGTGTGTGTGT 800
DB 340 AACTGACCTGCCAGGTGGTGTCTCTCAACAAATGCTTCCAGTGAAGGAGCTGTGTGTGTGTGTGTGTGTGT 399
QY 801 CCCTCAACTGGGTGGAGCACCAAGACAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 860
DB 400 CTGTCAACTGGGTGGAGCACCAAGACAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 459
QY 861 GGGCGGAGGCTGAGAAAGTACTGCGAGCTGAAGAACGCCCACTGGTGTGTGTGTGTGTGTGTGTGTGTGTGT 920
DB 460 GGGCGGAGGCTGAGAAAGTACTGCGAGCTGAAGAACGCCCACTGGTGTGTGTGTGTGTGTGTGTGTGTGTGT 519
```

QY 921 GGGAGGAGCAGAAATTTTGTCCAGAAATATCTAGGCTCCGCATACACCTGGATGGGCTCA 980
Db 520 GGGAGGAGCAGAAATTTTGTCCAGAAATATCTAGGCTCCGCATACACCTGGATGGGCTCA 579
QY 981 GTGACCTGAAGGAGCCTGGAAGTGGGTGGATGGGAACAGACTATGGACCGGCTTCCAGA 1040
Db 580 GTGACCTGAAGGAGCCTGGAAGTGGGTGGATGGGAACAGACTATGGACCGGCTTCCAGA 639
QY 1041 ACTGGAAGCCAGGCCAGCCAGCAGCTGGCAGGGGCAACGGCTGGGTGGAGCGGAGGACT 1100
Db 640 ACTGGAAGCCAGGCCAGCCAGCAGCTGGCAGGGGCAACGGCTGGGTGGAGCGGAGGACT 699
QY 1101 GTGCTCACTTCCATCCAGACGGCAGGTGGAATGACAGCTCTGCCAGAGGCCCTTACCCT 1160
Db 700 GTGCTCACTTCCATCCAGTCCGAGGTGGAATGACAGCTCTGCCAGAGGCCCTTACCCT 759
QY 1161 GGGTCTGCGAGGCTGGCTGGGTGAGCAGCCAGCAGGAGAGTCACTGAGTGGCTTTGGTG 1220
Db 760 GGGTCTGCGAGGCTGGCTGGGTGAGCAGCCAGCAGGAGAGTCACTGAGGTACCTTTGGTG 819
QY 1221 G 1221
Db 820 G 820

RESULT 11
US-10-152-319A-1777
; Sequence 1777, Application US/10152319A
; Publication No. US20040072160A1
; GENERAL INFORMATION:
; APPLICANT: Mendrick, Donna
; APPLICANT: Porter, Mark
; APPLICANT: Johnson, Kory
; APPLICANT: Higgins, Brandon
; APPLICANT: Castle, Arthur
; APPLICANT: Elashoff, Michael
; TITLE OF INVENTION: Molecular Toxicology Modeling
; FILE REFERENCE: 44921-5089-US
; CURRENT APPLICATION NUMBER: US/10/152,319A
; CURRENT FILING DATE: 2002-05-22
; PRIOR APPLICATION NUMBER: US 60/292,335
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/297,523
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,925
; PRIOR FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 60/303,810
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: US 60/303,807
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: US 60/303,808
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: US 60/315,047
; PRIOR FILING DATE: 2001-08-28
; PRIOR APPLICATION NUMBER: US 60/324,928
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US 60/330,867
; PRIOR FILING DATE: 2001-11-01
; PRIOR APPLICATION NUMBER: US 60/330,462
; PRIOR FILING DATE: 2001-10-22
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 2221
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1777
; LENGTH: 1358
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. NM_022393
US-10-152-319A-1777

Query Match 30.0%; Score 438; DB 17; Length 1358;
Best Local Similarity 68.1%; Pred. No. 7.9e-111;

Matches 676; Conservative 0; Mismatches 295; Indels 21; Gaps 4;
QY 210 TGCTCCATTTTCAGCTGTGCAACCTCAGAGCC---GTGTTGGCCCAAGACATGACAAGCA 266
Db 122 TGTCTCGGTTTCAGTTTCAGATAGCCCTAGAGCCCTTGTGTTTGTGCAAGCATGACATGG 181
QY 267 CGTATGAAATCTTCAGTACTTGGAGAAATAGGTGAAAGTCCAGGGGTTTAAAAATGGCC 326
Db 182 CATATGAAATCTTCAGAACTTGGGAGCGAAGAGAAAACCAAGAGGCTGGTAAAGC-- 239
QY 327 CACTTCCTCTCAGTCCCTCTGAGCGTCTCCGCTCTGGGCCCTTGGCATCTCTCTGCTGT 386
Db 240 ----TCCCTCCCACTTCCTGTGCAATATCTCTCTGGACCCACCTCTCTCTGTTCT 295
QY 387 CCCTGGGCTCCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 446
Db 296 CCCTGGGCTTAGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 355
QY 447 AATTTTCAGAGGACCTGCTGACCTCGAGAACAGATTTTTCAGCACTTCACTTCAACACTG 506
Db 356 AGTTAAGGAGGAGCCTAGAAACCTGAGAACCACTTTTAGACAACACCACTTCAACACCA 415
QY 507 TGGCGGAGATCCAGGCACTGACTTCCAGGGCAGAGCTTGGAGAGAAACGATAGCATCTC 566
Db 416 AGGCTGAACTACAGGCCCTTGGCTTCCAGGGGTGACAGCTTGCNAAACAGGAATCAATTCT 475
QY 567 TGAAGCTGAGCTGAGGGTTTCAAGCAGAGAAACGGCAGGAGGGGTATCTGAGCTCCAGG 626
Db 476 TGAAGCTGAGGTGATGATCATGGGCAGGAACTTGCAGGCGAGGCGGAGGCTTGGAGCC 535
QY 627 AACACACTACGAGAGGACACCTTAGGCCACTGTGCCCACTGTGCCCACTGTGTGTGTGTC 686
Db 536 AGGTGGCTTCTTGGAGAGCACAGTGGAGAGAAAGAGGACAGACTC-----TCAGAA 586
QY 687 CAGTTCACTTGAATGCTCTCTGAGTCCAGCAGCTGCTGCAAGACCTTGAAGAACTGA 746
Db 587 CAGATCTATCTGAAATTAACCGATCGTGTGCAACAGCTGGGGAAGGAGACTTGAAGAC 646
QY 747 CTGCGAGGTGGCTACTCTCAACAACTGCTCCACTGAAAGGACCTGTGCTGCCCGCTCA 806
Db 647 CATGTCAGCTGGCCAGCCTCAAGAAACAGGCTCAGC---AGTGGCTGTGCTGCCCTTC 703
QY 807 ACTGGGTGGAGCACCAAGACAGCTGCTACTGTGTTCTCTCACTTGGGATGTCTTGGGCG 866
Db 704 ACTGGATGGAGCATGAAGGCGAGCTGCTACTGTGTTCTCTCACTTGGGAAGCGCTGGC 763
QY 867 AGGCTGAGAACTGCTGCGAGCTGGAAGACGCCACCTGGTGGTCTATCACTTCCAGGAG 926
Db 764 AAGCTGACAGTACTGCGCAGCTGGAGAAATCTAAGCTAGTGGTGGTCAACTCCCTGG 823
QY 927 AGCAGAAATTTGTCCAGAAATATCTAGGCTCCGCATACACCTGGATGGGCTCAGTGACC 986
Db 824 AGCAGAAATTTTACAGACTCAGATGGGCACTGTGGTCACTTGGATGGCTTACCGACC 883
QY 987 CTGAAGGAGCCTTGAAGTGGGTGGATGGAAACAGACTATGCCACCGGCTTCCAGAACTGA 1046
Db 884 AAAATGGGCTTGGCGATGGGTGGAGCGGACCGACTATGAGAAAGGCTTTACGCACTGG 943
QY 1047 AGCCAGGCGAGCCAGACACTGCGCAGGGGCGAGGCTGGGTGGAGCGAGGACTGTGCTC 1106
Db 944 CCCCAGGAGCCAGATTAACCTGGTATGGACCGGCTGGGAGGAGGCGAGGACTGTGCAC 1003
QY 1107 ACTTCCATCCAGACGGCAGGTGGAAATGACAGCTCTGCCAGAGGCCCTTACCACCTGGGTCT 1166
Db 1004 ACTTCACTCAGATGGTCTGCTGGATGATGAGCTTTGCCAGAGGCCCTTACCGCTGGGTCT 1063
QY 1167 GCGAGGCTGGCTGGGTGAGACCGACCGAGGAG 1198
Db 1064 GTGAGATGAAGCTGGCCAAAGGACGACTAGGAG 1095

RESULT 12
US-09-880-107-2230

; Sequence 2230, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2230
; LENGTH: 1277
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 M10058
US-09-880-107-2230

Query Match 26.2%; Score 382; DB 9; Length 1277;
Best Local Similarity 66.1%; Pred. No. 3e-95;
Matches 646; Conservative 0; Mismatches 245; Indels 87; Gaps 3;

QY 224 CTGTGACCACTCAGAGCCCTGTGGCCCAAGCATGACAGGACGATGAAATTCAG 283
DB 140 CTGAGCAATCCCAGGTCAGCGCCATATCATGACCAAGAGTAAAGACCTTCAG 199

QY 284 TACTTGGAGAAATAGGTGAAAGTCCAG---GGGTTTAAATAATGGCCACTTCTCTCCAG 340
DB 200 CATCTGGACAATGAGGAGAGTGACCATCAGCTCAGAAAAGGCCACCTCTCCAG 259

QY 341 TCCTCTCTGAGCGTCTCCGCTCTGGGCGCTGCGCATCTCTGTGTCTCTCTGGGCTCGGC 400
DB 260 CCCCTCTGAGCGTCTCTGCTCGGACCTCGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 319

QY 401 CTGCTCTGCTGTCATCTCTGT 460
DB 320 CTCTCTCTCTGT 379

QY 461 CTGCTGACCTCTGAGAGACAGATTTAGCAACTTCACCTCAAACTGTGGCGGAGATCCAG 520
DB 380 CTGGGGCTGTGAGAGAGAGTTCAGCACTTCACAGGAGACGGAGGCCAGGTCAAG 439

QY 521 GCACTGACTTTCCAGGCGACAGCTTGGAGAAACGATAGCATCTCTGAAAGCTGAGGTG 580
DB 440 GGCTTGAGCACCAGGGAGGCAATGTGGGAAGAAAGATGAAGTCGCTAGAGTCCAGCTG 499

QY 581 GAGGTTTCAAGCAGGAAACGGCAGGAGGAGGTATCTGAGCTCCAGGAAACACTACGAC 640
DB 500 GAGAAACAGCAGAGGACCTGAGTG-----524

QY 641 AAGGCACACCTAGGCCACTGTCCCACCTGCTGTCCTGTCGTCGTCGTCGTCGTCGTCGTCG 700
DB 525 -----AAGATCACTCAGC 538

QY 701 ATGCTCTGCGAGTCCAGAGCTGGTCAAGACCTGAAGAAACTGACTCTCCAGGTGGCT 760
DB 539 CTGCTGTCTCAGCTGAGAGAGTTCGTGTCTGACCTGGGAGCCTGAGCTGTGAGATGGCG 598

QY 761 ACTCTCAACAAATGCTCTCACTGAAGGAGACCTGTGCGCGCTCAACTGCGGTGGAGCAC 820
DB 599 GCGCTCCAGGCGCAATGGCT---CAGAAAGGACCTGTGCTGCGCGCTCAACTGCGGTGGAGCAC 655

QY 821 CAAGACAGCTGCTACTGCTTCTCTCACTCTGGGATGTCCTGGGCGGAGGCTGAGAGATAC 880
DB 656 GAGGCGAGCTGCTACTGCTTCTCTCTGCTCCGGGAAGGCGCTGGGCTGACGCCGCAACTAC 715

QY 881 TGCAGCTGAAGAAACGCCCACTGGTGGTTCATCACTCCAGGAGGAGCAATTTTGTCT 940
DB 716 TGCCTGCTGAGAGACGCGCACTGGTGGTTCATCGCTCTGGGAGGAGCAAAATTTGTCT 775

QY 941 CAGAAATATCTAGGCTCCGCATACACTGTGATGGCTCAGTGACCCCTGAGGAGGCTGG 1000
DB 776 CAGCAACATAGGCTCTGTGAACACTGTGATGGGCTCCACCAAAAGGCGCTGG 835

QY 1001 AAGTGGGTGAGAAACAGACTATGCGACCGCTTCAGAACTTGAAGCCAGGCGCAACCA 1060
DB 836 AAGTGGGTGAGCGGAGCGGACTACGAGACGGGCTTCAAGAACTGGAGGCGGAGCAGCG 895

QY 1061 GACGACTGGCAGGCGGACCGGCTGGGTGGAGGAGGAGTGTCTCACTTCCATCCAGAC 1120
DB 896 GACGACTGGTACGCGCCACCGGCTCGGAGGAGGAGGAGTGTGTGCCCACTTCAACGACGAC 955

QY 1121 GGCAGGTGGAATGACGACGCTCTGCCAGAGGCGCTTACACTGGGCTCTGCCAGGCTGGCCTG 1180
DB 956 GGCCTGCTGNAACGACGACGCTCTGCCAGAGGCGCTTACCGTGGGTCTGCCGAGACAGAGCTG 1015

QY 1181 GGTCCAGACCCAGGAG 1198
DB 1016 GACAAGGCCAGCCAGGAG 1033

RESULT 13
US-10-305-720-1186
; Sequence 1186, Application US/10305720
; Publication No. US20040010136A1
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice K.; Seilhamer, Jeffrey J.
; TITLE OF INVENTION: Composition for the Detection of Signaling Pathway Gene Expression
; FILE REFERENCE: PA-0002-1 CON
; CURRENT APPLICATION NUMBER: US/10/305,720
; CURRENT FILING DATE: 2002-11-26
; PRIOR APPLICATION NUMBER: 09/016,434
; PRIOR FILING DATE: 1998-01-30
; NUMBER OF SEQ ID NOS: 1490
; SOFTWARE: PERL Program
; SEQ ID NO 1186
; LENGTH: 1277
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: GenBank ID No. US20040010136A1 g179078
US-10-305-720-1186

Query Match 26.2%; Score 382; DB 17; Length 1277;
Best Local Similarity 66.1%; Pred. No. 3e-95;
Matches 646; Conservative 0; Mismatches 245; Indels 87; Gaps 3;

QY 224 CTGTGACCACTCAGAGCCCTGTGGCCCAAGCATGACAGGACGATGAAATTCAG 283
DB 140 CTGAGCAATCCCAGGTCAGCGCCATATCATGACCAAGAGTAAAGACCTTCAG 199

QY 284 TACTTGGAGAAATAGGTGAAAGTCCAG---GGGTTTAAATAATGGCCACTTCTCTCCAG 340
DB 200 CATCTGGACAATGAGGAGAGTGACCATCAGCTCAGAAAAGGCCACCTCTCCAG 259

QY 341 TCCTCTCTGAGCGTCTCCGCTCTGGGCGCTGCGCATCTCTGTGTCTCTCTGGGCTCGGC 400
DB 260 CCCCTCTGAGCGTCTCTGCTCGGACCTCGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 319

QY 401 CTGCTCTGCTGTCATCTCTGT 460
DB 320 CTCTCTCTCTGT 379

QY 461 CTGCTGACCTCTGAGAGACAGATTTAGCAACTTCACCTCAAACTGTGGCGGAGATCCAG 520
DB 380 CTGGGGCTGTGAGAGAGAGTTCAGCACTTCACAGGAGACGGAGGCCAGGTCAAG 439

QY 521 GCACTGACTTTCCAGGCGACAGCTTGGAGAAACGATAGCATCTCTGAAAGCTGAGGTG 580
DB 440 GGCTTGAGCACCAGGGAGGCAATGTGGGAAGAAAGATGAAGTCGCTAGAGTCCAGCTG 499

QY 581 GAGGTTTCAAGCAGGAAACGGCAGGAGGAGGTATCTGAGCTCCAGGAAACACTACGAC 640
DB 500 GAGAAACAGCAGAGGACCTGAGTG-----524

QY 641 AAGGCACACCTAGGCCACTGTCCCACCTGCTGTCCTGTCGTCGTCGTCGTCGTCGTCGTCG 700
DB 525 -----AAGATCACTCAGC 538

QY 701 ATGCTCTGCGAGTCCAGAGCTGGTCAAGACCTGAAGAAACTGACTCTCCAGGTGGCT 760
DB 539 CTGCTGTCTCAGCTGAGAGAGTTCGTGTCTGACCTGGGAGCCTGAGCTGTGAGATGGCG 598

QY 761 ACTCTCAACAAATGCTCTCACTGAAGGAGACCTGTGCGCGCTCAACTGCGGTGGAGCAC 820
DB 599 GCGCTCCAGGCGCAATGGCT---CAGAAAGGACCTGTGCTGCGCGCTCAACTGCGGTGGAGCAC 655

QY 821 CAAGACAGCTGCTACTGCTTCTCTCACTCTGGGATGTCCTGGGCGGAGGCTGAGAGATAC 880
DB 656 GAGGCGAGCTGCTACTGCTTCTCTCTGCTCCGGGAAGGCGCTGGGCTGACGCCGCAACTAC 715

Db 440 GGCTTTCAGCACCCAGGAGGCAATGTGGGAAGAAAGATGAAGTCGTAGTCCAGCTG 499
Qy 581 GAGGGTTTCAAGCAGGAACGGCAGGAGGGGTATCTGAGCTCCAGGACACACTAGCGAG 640
Db 500 GAGAAACAGCAGGAAGACCTGAGTG----- 524
Qy 641 AAGGCACACTAGGCCACTGTGTCCTCCACTGCCATCTGTGTGTGCCAGTTTCATTCTGAA 700
Db 525 -----AAGATCACTCCAGC 538
Qy 701 ATGCTCTCGAGTCCAGCAGCTGGTGGCAAGACCTGAAGAACTGACCTGCCAGTGCGT 760
Db 539 CTGCTCTCCAGCTGAAGCAGTTCGTGTCTGACCTGCGGAGCCTGAGCTGTCCAGATGGCG 598
Qy 761 ACTCTCAACAAATGCTCCACTGAAGGAGCTGCTGCCCGCTCAACTGGGTGGAGCAC 820
Db 599 GCGCTCCAGGCAATGGCT---CAGAAAGGACCTGTGCTGCCGGTCAACTGGGTGGAGCAC 655
Qy 821 CAAGACAGCTGTACTGTGTTCTCTCACTCTGGGATGTCTCTGGGCGGAGGCTCAGAACTAC 880
Db 656 GAGCGCAGCTGTACTGTGTTCTCTGCTCCGGGAGGCTTGGGCTGACGCCGACAACTAC 715
Qy 881 TGCAGCTGAAGAACCCCACTGTGTGTCTCACTCAGGAGGAGCAGAAATTTGTG 940
Db 716 TGCCTGCTGGAGGACGCGCAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 775
Qy 941 CAGAAATATCTAGGCTCCGCTACACCTGATGAGGCTTCCAGTGTGAGGAGGCTTGAAGGAGCTGG 1000
Db 776 CAGCACCACATAGGCGCTGTGTGAACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 835
Qy 1001 AAGTGGGTGGATGGAACAGACTATGCGACCGGCTTCCAGAACTGGAAGCGCCAGCGCA 1060
Db 836 AAGTGGGTGACGGGACGACTACGAGACGGGCTTCAAGAACTGAGGCGCGAGCAGCG 895
Qy 1061 GACGACTGCGAGGGGACCGGGCTGGGTGGAGCGGAGGACTGTGTCTCACTTCCATCCAGAC 1120
Db 896 GACGACTGTGTGCGCCACCGGCTCGGAGGAGCGAGGACTGTGTGTGTGTGTGTGTGTGTGTGT 955
Qy 1121 GCGAGGTGGAATCAGCACTCTGCGAGGCGCTTACCACTGGGTCTGCGAGGCTGCGCTG 1180
Db 956 GCGCGGTGGAAGCAGCACTCTGCGAGGCGCTTACCGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1015
Qy 1181 GGTGACAGCAGCAGGAG 1198
Db 1016 GACAGGCGCAGCAGGAG 1033

RESULT 14
US-10-236-392-9
; Sequence 9, Application US/10236392
; Publication No. US20040067490A1
; GENERAL INFORMATION:
; APPLICANT: Anderson, David W
; APPLICANT: Boldog, Ferenc L
; APPLICANT: Burgess, Catherine, E
; APPLICANT: Casman, Stacie J
; APPLICANT: Catterton, Elina
; APPLICANT: Chapoval, Andrei
; APPLICANT: Crabtree, Julie
; APPLICANT: Edinger, Shlomit, R
; APPLICANT: Ellerman, Karen
; APPLICANT: Gerlach, Valerie
; APPLICANT: Gorman, Linda
; APPLICANT: Grosse, William M
; APPLICANT: Gusev, Vladimir
; APPLICANT: Kekuda, Ramesh
; APPLICANT: LaRoche, William J
; APPLICANT: Li, Li
; APPLICANT: MacDougall, John R
; APPLICANT: Malyankar, Uriel M
; APPLICANT: Miller, Charles E
; APPLICANT: Millet, Isabelle

; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Patturajan, Meera
; APPLICANT: Pena, Carol A
; APPLICANT: Peyman, John A
; APPLICANT: Rastelli, Luca
; APPLICANT: Reiger, Daniel K
; APPLICANT: Rothenberg, Mark E
; APPLICANT: Shenoy, Suresh
; APPLICANT: Shimkets, Richard A
; APPLICANT: Smithson, Glenda
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-442A
; CURRENT APPLICATION NUMBER: US/10/236,392
; CURRENT FILING DATE: 2002-09-06
; PRIOR APPLICATION NUMBER: US09/540,763
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: US60/390,155
; PRIOR FILING DATE: 2002-06-19
; PRIOR APPLICATION NUMBER: US09/635,949
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: US60/318,765
; PRIOR FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: US60/357,303
; PRIOR FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: US60/367,753
; PRIOR FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER: US60/369,479
; PRIOR FILING DATE: 2002-04-02
; PRIOR APPLICATION NUMBER: US09/659,634
; PRIOR FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: US60/318,120
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: US60/318,130
; PRIOR FILING DATE: 2001-09-07
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 794
; SOFTWARE: Custom
; SEQ ID NO 9
; LENGTH: 1055
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (121)..(982)
US-10-236-392-9

Query Match 22.0%; Score 321.4; DB 17; Length 1055;
Best Local Similarity 62.1%; Pred. No. 2e-78;
Matches 591; Conservative 0; Mismatches 276; Indels 84; Gaps 2;
Qy 248 GGCCCAAGCATGACAAAGGACGCTATGAAAACTTCCAGTACTTGGAGAAATAAGGTGAAAGTC 307
Db 112 GGCCCCATCTGGCCAGGAGCTTTCAAGATATCCAGCAGCTGAGCTGGAGGAAAATGAC 171
Qy 308 CAGGGGTTTAAAAATGGGCCACTTCTCTTCCAGTCCCTCTGCGAGCGGTCTCCGCTCTGGG 367
Db 172 CATCTTTTCCATCAAGGGCCACCTCTCTGCCCAGCCCTGGCAGCAGCGTCTCTGCTCCATG 231
Qy 368 CCTGGCATCTCTGCTGTCTCTCTGGGCTCGGCTGTCTGTCTGTCTGTCTGTCTGTCTGTG 427
Db 232 GTCTGCTTCACTCTGCTTGTGCTTCAACATCTCTGCTGTCTGTGCTGTCTGTCTGTG 291
Qy 428 GTTGGATTCCAAAATTTCCAAAATTTTCAGAGGAGCCTGGTGACCTTGAGAACAGATTTTACG 487
Db 292 ACTGGTCCCAAGTGCACAGCTGCAAGCCAGCTGCGAGCCTTGAGGAAGCTTTTCAGC 351
Qy 488 AACTTCACCTCAAAACACTGTGTGGCGAGATCCAGGCACCTGACTTCCAGGGCAGCAGCTTG 547
Db 352 AACTTCTCTCGAGCACCTTGCAGGAGGTCCAGGCAATCAGCACCCACGAGGAGCGCTG 411
Qy 548 GAAGAAACGATAGCATCTCTGAAAGCTGAGTGGAGGGTTTCAAGCAGGAAACGGCAGCA 607
Db 412 GGTGACAAAGATCACATCTCTAGGAGCCAAGCTCTAGGAGAAACAGCAGCAGGACCTGAAAG-- 469

```
QY 608 GGGGTATCTGAGCTCCAGGAACACACTACGAGAGGACACACTAGGCGCACTGTCCCCAC 667
Db 470 -----
QY 668 TGCCCATCTGTGTGTCTCCAGTTTCATTTCTGAATGCTCTCTGAGGTCCAGCAGCTGGTG 727
Db 470 -----CAGATCAGATGCTCTCTTCCATCTGAAGCACTTCCCC 510
QY 728 CAAGACCTGAAGAACTGACTGTCAGGTGGCTACTCTCAACAACAAATGCTCCACAGAA 787
Db 511 GTGGACCTGGGCTGCTGGGCTGTCAGATGAGCTCTCCACAGCAACGGCTCC---CAA 567
QY 788 GGGACCTGCTGCCCGCTCACTGGGTGGAGCACCAAGACAGCTGCTACTGTTCTCTCAC 847
Db 568 AGGACCTGCTGCCCGCTCACTGGGTGGAGCACCAAGGCACTGCTACTGTTCTCTCAC 627
QY 848 TCTGGGATGTCTGGGCGGAGGCTGAGAAGTACTGCCAGCTGAGAAGCGCCCACTGGTG 907
Db 628 TCCGGGAGGCTGGGCTGAGGCGGAGAGTACTGCCAGCTGGAGAACGACACCTGGTG 687
QY 908 GTCACTAACTCCAGGAGGAGAGAAATTTGTCCAGAAATATCTAGGCTCCGATACACC 967
Db 688 GTCACTAACTCTGGGAGGAGAGAAATTTGTACACACACAGCAACCCCTTCAATACC 747
QY 968 TGGATGGCTCTAGTGACCTGAGGAGCTGGAAGCTGGGATGGGTGGATGGAAACAGACTATGCG 1027
Db 748 TGGATAGGCTCTAGGAGAGCTGATGGCTCTTGGAAATGGGTGGATGGACAGACTATAGG 807
QY 1028 ACCGGCTTCCAGAACTCGAAGCCAGGCGCAGCAGCACTGGCAGGGGCGACGGGCTGGGT 1087
Db 808 CACAACTACAGAACTGGGCTGTCTCAGCAGGATTAATGGACAGGCGCAGAGCTGGGT 867
QY 1088 GGAGGCGAGGACTGTGCTACTTCCATCCAGACGGCAGGTGGAAATGACAGCTGTGCCAG 1147
Db 868 GGAAGTGAAGACTGTGTTGAAGTCCAGCCGATGGCGCTGGAAACGATGACTTCTGCTG 927
QY 1148 AGGCCCTACCACTGGGTCTGGAGGCTGGCTGGGTGAGACCGACCGAGGAG 1198
Db 928 CAGGTGTACCGCTGGGTGTGTGAGAAAAGCGGAATGCCACCGCGAGGTG 978
```

RESULT 15

```
US-09-880-107-3731
; Sequence 3731, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3731
; LENGTH: 1300
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 X55283
US-09-880-107-3731
```

```
Query Match 22.0%; Score 321.4; DB 9; Length 1300;
Best Local Similarity 62.1%; Pred. No. 2.2e-78;
Matches 591; Conservative 0; Mismatches 276; Indels 84; Gaps 2;
```

Search completed: May 28, 2005, 02:07:55
Job time : 928 secs

```
QY 248 GGGGCAAGCATGACAAAGCGTATGAAACTTCCAGTACTTGGAGTAAGCTGAAAGTC 307
Db 182 GGGCCCATCATGCGCCAGGACTTTCAAGATATCAGCAGCTGAGCTCGGAGGAAATGAC 241
QY 308 CAGGGGTTTTAAATGGGCCACTTCTCTCAGTCCCTCTGAGCGCTCTCCGCTCTGGG 367
Db 242 CATCTTTCCATCAAGGGCCACTCTCTGCCAGCCCTGGCAGCGTCTCTGCTCCATG 301
QY 368 CCCTGCCATCTCTGCTGCTGGGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 427
Db 302 GTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 361
QY 428 GTTGGATTTCCAAATTTCCAAATTTCCAAATTTCCAAATTTCCAAATTTCCAAATTTCC 487
Db 362 ACTGGGTCCTCCAAAGTGCACAGCTGCAAGCGGAGCTGGGAGCTTGAAGAGCTTTTCAG 421
QY 488 AACTTTCACCTTCAAACTCTGTGGCGGAGATCCAGGCACTGACTTTCACAGGCGCAGCTTG 547
Db 422 AACTTCTCTCGAGCACTCTGAGGAGTCCAGGCAATCAGCACCCCAACGAGGCGCGTG 481
QY 548 GAAGAAACGATAGCATCTCTGAAAGCTGAGGTGGAGGTTTCAAGCAGGAAACGCGAGGCA 607
Db 482 GGTGCAAGATCATCTCTTAGGAGCCAGCTGGAGAAACAGCAGCAGGACCTGAAAG-- 539
QY 608 GGGGTATCTGAGCTCCAGGAAACACTAGCAGAGGACACACTAGCAGAGGACACACTAGGCC 667
Db 540 -----
QY 668 TGCCCATCTGTGTGTGTCTCCAGTTTCATTTCTGAATGCTCTCTGAGGTCCAGCAGCTGGTG 727
Db 540 -----CAGATCAGATGCTCTCTTCCATCTGAAGCACTTCCCC 580
QY 728 CAAGACTGAAAGAACTGACTGCGAGGTGGCTACTCTCAACAACAAATGCTCTCAGCTGAA 787
Db 581 GTGCACTGCGCTCTGCTGGCTGCGAGTGGAGTCTCTCCACAGCAACGGCTCC---CAA 637
QY 788 GGGACCTGCTGCGCCCTCACTGGGTGGAGCACCAAGACAGCTGCTACTGCTTCTCTCAC 847
Db 638 AGGACCTGCTGCGCCCTCACTGGGTGGAGCACCAAGGAGCTGCTACTGCTTCTCTCAC 697
QY 848 TCTGGGATGTCTGGGCGGAGGCTGAGAAGTACTTGGCAGCTGAGAAGCGCCCACTGGTG 907
Db 698 TCCGGGAGGCTGGGCTGAGGCGGAGAGTACTTCCAGCTGGAGAACGACACACTGGTG 757
QY 908 GTCATCAACTCCAGGAGGAGAGAAATTTGTCCAGAAATATCTAGGCTCTCGCATACACC 967
Db 758 GTCATCAACTCTCTGGGAGGAGAGAAATTTCAATTTGTAACAACACAGCAACCCCTTCAATACC 817
QY 968 TGGATGGGCTCAGTGACCTGAGGAGCTGGAAGTGGGTGGATGGAAACAGACTATGCG 1027
Db 818 TGGATAGGCTCTCAGGAGAGTGTGCTTGGAAATGGGTGGATGGACAGACTATAGG 877
QY 1028 ACCGGCTTCCAGAACTGGAAGCCAGGCGCAGCAGCACTGCGAGGGGCGCAGGCTGGGT 1087
Db 878 CACAACTACAAAGAACTGGGCTGTCTCAGCAGCAGATTAATGGCAGGCGCAGGCTGGGT 937
QY 1088 GGAGGCGAGGACTGTGCTCTACTTCCATCCAGACGGCAGGTGGAAATGACAGCTCTGCCAG 1147
Db 938 GGAAGTGAAGACTGTGTTGAAGTCCAGCGGATGGCGCTGGAACGATGACTTCTGCTG 997
QY 1148 AGGCCCTACCACTGGGTCTCGAGGCTGGGCTGAGTGGGTGAGACCGACCGAGGAG 1198
Db 998 CAGGTGTACCGCTGGGTGTGTGAGAAAAGCGGAATGCCACCGCGAGGTG 1048
```

This page blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 27, 2005, 12:44:45 ; Search time 65 Seconds
(without alignments)
2489.493 Million cell updates/sec

Title: US-10-829-107-4
Perfect score: 1738
Sequence: 1 MTRTYENFQYLENKNKVGQF.....QRPYHWCEAGLQTSQESH 316

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Uniprot 03:*

1: uniprot_sprot:*

2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1730	99.5	316	2 Q8IUN9	Q8IUN9 homo sapien
2	1544	88.8	292	2 Q14538	Q14538 homo sapien
3	998.5	57.8	256	2 Q6PIW3	Q6PIW3 homo sapien
4	918.5	52.8	290	1 LECH HUMAN	P07306 homo sapien
5	887	51.0	306	1 NMGL RAT	P49301 rattus norv
6	877	50.5	284	2 Q91V84	Q91V84 mus musculu
7	874	50.3	283	1 LECH RAT	P02706 rattus norv
8	870	50.1	332	2 Q8JZNI	Q8JZNI mus musculu
9	866	49.8	283	1 LECH MOUSE	P34927 mus musculu
10	805	46.3	304	2 Q91VT3	Q91VT3 mus musculu
11	804	46.3	304	1 NMGL MOUSE	P49300 mus musculu
12	791	45.5	361	2 Q8BUD5	Q8BUD5 mus musculu
13	767	44.1	306	2 Q7Z4G9	Q7Z4G9 homo sapien
14	754.5	43.4	311	1 LECI HUMAN	P07307 homo sapien
15	671.5	38.6	301	1 LECI MOUSE	P24721 mus musculu
16	633.5	36.4	301	1 LECI RAT	P08290 rattus norv
17	619.5	35.6	142	2 Q6DUK3	Q6DUK3 marmota mon
18	363	20.9	144	2 Q6DD06	Q6DD06 xenopus lae
19	344	19.8	98	2 Q6DUK2	Q6DUK2 marmota mon
20	331.5	19.1	381	1 C209 HYLSY	Q8HY02 hylobates s
21	326.5	18.8	550	1 KUCR RAT	P10716 rattus norv
22	325.5	18.7	381	1 C209 MAGNE	Q951C6 macaca neme
23	322.5	18.6	404	1 C209 PONPY	Q8HY00 pongo pygma
24	322	18.5	399	1 C209L HYLCO	Q8HY10 hylobates c
25	320.5	18.4	381	1 C209 CERAE	P60883 cercopithec
26	320.5	18.4	422	1 C209L HYLSY	Q8HY11 hylobates s
27	320	18.4	404	1 C209 MACMU	Q95J96 macaca mula
28	319	18.4	358	1 C209 PAPHA	Q8HY04 papio hamad
29	319	18.4	399	1 C209L HYLLA	Q8HY12 hylobates l
30	316.5	18.2	353	2 Q69F40	Q8HY06 homo sapien
31	316.5	18.2	376	1 C209L GORGO	Q8HY06 gorilla gor

32	316	18.2	404	1 C209 HUMAN	Q8NNX6 homo sapien
33	313.5	18.0	427	1 C209_GORGO	Q8HX28 gorilla gor
34	313	18.0	404	1 C209_HYLCO	Q8HY01 hylobates c
35	313	18.0	742	2 Q6P9F2	Q6P9F2 homo sapien
36	313	18.0	742	2 Q9BYH7	Q9BYH7 homo sapien
37	312	18.0	399	1 C209L HUMAN	Q8H2X3 homo sapien
38	312	18.0	427	1 C209_PANTR	Q8HX27 pan troglod
39	311.5	17.9	548	2 Q8BLZ8	Q8BLZ8 mus musculu
40	311	17.9	417	2 Q8TCR2	Q8TCR2 homo sapien
41	311	17.9	742	2 Q8WZA4	Q8WZA4 homo sapien
42	310.5	17.9	450	1 C209 HYLLA	Q8HY03 hylobates l
43	309.5	17.8	548	1 KUCR MOUSE	P70194 mus musculu
44	308	17.7	445	1 C209L_PANTR	Q8HY00 pan troglod
45	304.5	17.5	260	2 Q7TSU6	Q7TSU6 mus musculu

ALIGNMENTS

RESULT 1

Q8IUN9	ID	Q8IUN9	PRELIMINARY;	PRT;	316 AA.
AC	Q8IUN9;				
DT	01-MAR-2003	(T-EMBLrel. 23, Created)			
DT	01-MAR-2003	(T-EMBLrel. 23, Last sequence update)			
DT	01-MAR-2004	(T-EMBLrel. 26, Last annotation update)			
DE	C-type lectin, superfamily member 13, isoform 1.				
GN	Name=CLECSF14;				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Brain;				
RX	MDLLINE=223388257; PubMed=12477932; DOI=10.1073/pnas.242603899;				
RA	Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,				
RA	Klausner R.D., Collins F.S., Wagner K.H., Shenmen C.M., Schuler G.D.,				
RA	Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,				
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heieh F.,				
RA	Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,				
RA	Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,				
RA	Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,				
RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,				
RA	Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,				
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,				
RA	Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,				
RA	Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,				
RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,				
RA	Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,				
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,				
RA	Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,				
RA	Jones S.J., Marra M.A.;				
RT	"Generation and initial analysis of more than 15,000 full-length human				
RT	and mouse cDNA sequences."				
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).				
[2]					
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Brain;				
RA	Strausberg R.;				
RL	Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; BC039011; AAH39011.1; -				
DR	HSSP; P07306; IDV8.				
DR	Genew; HGNC:16916; CLECSF14.				
DR	GO; GO:0016020; C-membrane; IEA.				
DR	GO; GO:0005529; F-sugar binding; IEA.				
DR	InterPro; IPR002353; AntifreezeZell.				
DR	InterPro; IPR001304; Lectin_C.				
DR	InterPro; IPR005640; Lectin_N.				
DR	Pfam; PF00059; Lectin_C; 1.				
DR	Pfam; PF03954; Lectin_N; 1.				
DR	PRINTS; PR00356; ANTI-FREEZEII.				
DR	SMART; SM00034; CLECT; 1.				


```
DR PROSITE, P800615; C_TYPE LECTIN_1; 1.
DR PROSITE, P850041; C_TYPE LECTIN_2; 1.
KW Lectin.
SQ SEQUENCE 316 AA; 35446 MW; D3B7193E2E1F58AF CRC64;

Query Match 99.5%; Score 1730; DB 2; Length 316;
Best Local Similarity 99.7%; Pred. No. 2.2e-136;
Matches 315; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MTRTYENFOYLENKVKVQGFKNGLPLQSLQRLRSGPCHLLSLGLGLLLVLCVVG 60
DB 1 MTRTYENFOYLENKVKVQGFKNGLPLQSLQRLRSGPCHLLSLGLGLLLVLCVVG 60

QY 61 QNSKFQDRLVLTDFSNFTSNVAIEIQAITSQSSLEETIASLKAEEVGFQKQRAQVS 120
DB 61 QNSKFQDRLVLTDFSNFTSNVAIEIQAITSQSSLEETIASLKAEEVGFQKQRAQVS 120

QY 121 ELQEHHTQKAHLGHCHPCPSVCPVHSEMLLRVQQLVQDLKLTQCVATLNNN---ASTE 177
DB 118 -----VHSEMLLRVQQLVQDLKLTQCVATLNNNNGEEASTE 153

QY 178 GTCCPVNVVHSDSCYWFSGMSWAEAEKYCOLKNAHLVWINSREONFVKYLGSAYT 237
DB 154 GTCCPVNVVHSDSCYWFSGMSWAEAEKYCOLKNAHLVWINSREONFVKYLGSAYT 213

QY 238 WMGLSDPEGAWKWDGTDYATGFQNWKPQDDWQHGGLGGEDCAHFHPDGRWDDVCQ 297
DB 214 WMGLSDPEGAWKWDGTDYATGFQNWKPQDDWQHGGLGGEDCAHFHPDGRWDDVCQ 273

QY 298 RPYHWVCEAGLGQTSQESH 316
DB 274 RPYHWVCEAGLGQTSQESH 292

RESULT 2
Q14538 ID Q14538 PRELIMINARY; PRT; 292 AA.
AC Q14538;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE Macrophage lectin 2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96108853; PubMed=8598452;
RA Suzuki N., Yamamoto K., Toyoshima S., Osawa T., Irimura T.;
RT "Molecular cloning and expression of cDNA encoding human macrophage c-
type lectin: Its unique carbohydrate binding specificity for Th
antigen.";
RL J. Immunol. 156:128-135(1996).
DR EMBL; D50532; BAA09101.1; -.
DR HSSP; P07306; 1DV8.
DR GO; GO:0005886; C:plasma membrane; TAS.
DR GO; GO:0005529; F:sugar binding; TAS.
DR InterPro; IPR002353; AntifreezeII.
DR InterPro; IPR001304; Lectin_C.
DR Pfam; PF00059; Lectin_N; 1.
DR Pfam; PF03954; Lectin_C; 1.
DR PRINTS; PR00356; ANTI-FREEZEII.
DR SMART; SM00034; CLECT; 1.
DR PROSITE; PS00615; C_TYPE LECTIN_1; 1.
DR PROSITE; PS50041; C_TYPE LECTIN_2; 1.
KW Lectin.
SQ SEQUENCE 292 AA; 32937 MW; 54743389EBC80919 CRC64;

Query Match 88.8%; Score 1544; DB 2; Length 292;
Best Local Similarity 90.6%; Pred. No. 7.3e-121;
Matches 289; Conservative 0; Mismatches 0; Indels 30; Gaps 2;
```

```
QY 1 MTRTYENFOYLENKVKVQGFKNGLPLQSLQRLRSGPCHLLSLGLGLLLVLCVVG 60
DB 1 MTRTYENFOYLENKVKVQGFKNGLPLQSLQRLRSGPCHLLSLGLGLLLVLCVVG 60

QY 61 QNSKFQDRLVLTDFSNFTSNVAIEIQAITSQSSLEETIASLKAEEVGFQKQRAQVS 120
DB 61 QNSKFQDRLVLTDFSNFTSNVAIEIQAITSQSSLEETIASLKAEEVGFQKQRAQVS 117

QY 121 ELQEHHTQKAHLGHCHPCPSVCPVHSEMLLRVQQLVQDLKLTQCVATLNNN---ASTE 177
DB 118 -----VHSEMLLRVQQLVQDLKLTQCVATLNNNNGEEASTE 153

QY 178 GTCCPVNVVHSDSCYWFSGMSWAEAEKYCOLKNAHLVWINSREONFVKYLGSAYT 237
DB 154 GTCCPVNVVHSDSCYWFSGMSWAEAEKYCOLKNAHLVWINSREONFVKYLGSAYT 213

QY 238 WMGLSDPEGAWKWDGTDYATGFQNWKPQDDWQHGGLGGEDCAHFHPDGRWDDVCQ 297
DB 214 WMGLSDPEGAWKWDGTDYATGFQNWKPQDDWQHGGLGGEDCAHFHPDGRWDDVCQ 273

QY 298 RPYHWVCEAGLGQTSQESH 316
DB 274 RPYHWVCEAGLGQTSQESH 292

RESULT 3
Q6PIW3 ID Q6PIW3 PRELIMINARY; PRT; 256 AA.
AC Q6PIW3;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE CLCSF14 protein.
GN Name=CLCSF14;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E.,
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC027858; AAH27858.1; -.
DR HSSP; Q9H8F0; 1K9J.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005529; F:sugar binding; IEA.
DR InterPro; IPR001304; Lectin_C.
```

DR InterPro: IPR005640; lectin_N.
 DR Pfam; PF00059; Lectin_C; 1.
 DR Pfam; PF03954; Lectin_N; 1.
 SQ SEQUENCE 256 AA; 28958 MW; C2C1A576F369647D CRC64;
 Query Match 57.5%; Score 998.5; DB 2; Length 256;
 Best Local Similarity 87.6%; Pred. No. 2.7e-75;
 Matches 197; Conservative 1; Mismatches 0; Indels 27; Gaps 1;
 QY 1 MTRTYENFOYLENKVKVQGFKNGLPLQSLQLRLRSGPCHLLSLGLGLLLVVICVVG 60
 DB 1 MTRTYENFOYLENKVKVQGFKNGLPLQSLQLRLRSGPCHLLSLGLGLLLVVICVVG 60
 QY 61 QNSKFORDLVLTDFSNFTNTVAETQALTSQSSLEETIASLKAEVEGFKQERQAGVS 120
 DB 61 QNSKFORDLVLTDFSNFTNTVAETQALTSQSSLEETIASLKAEVEGFKQERQAGVS 117
 QY 121 ELQEHHTQKAHLGHCPHPCVCPVHSEMLLRVQQLVQDLKLTTCVATLNNASTEGTC 180
 DB 118 -----VHSEMLLRVQQLVQDLKLTTCVATLNNASTEGTC 153
 QY 181 CPVNVWEHQDSCYWFHSGMSMAEAEKYCOLKNAHLVWVINSREQ 225
 DB 154 CPVNVWEHQDSCYWFHSGMSMAEAEKYCOLKNAHLVWVINSREQ 198
 RESULT 4
 LECH HUMAN
 ID LECH HUMAN STANDARD; PRT; 290 AA.
 AC P07306;
 DT 01-APR-1988 (Rel. 07, Created)
 DT 01-APR-1988 (Rel. 07, Last sequence update)
 DT 25-OCT-2004 (Rel. 45, Last annotation update)
 DE Asialoglycoprotein receptor 1 (Hepatic lectin H1) (ASGPR) (ASGP-R).
 GN Name=ASGPR1;
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP MEDLINE=85130911; PubMed=2982798;
 RX Spiess M., Schwartz A.B., Lodish H.F.;
 RA "Sequence of human asialoglycoprotein receptor cDNA. An internal
 RT signal sequence for membrane insertion."
 RL J. Biol. Chem. 260:1979-1982 (1985).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=86079574; PubMed=3753585; DOI=10.1016/0092-8674(86)90496-4;
 RA Spiess M., Lodish F.;
 RT "An internal signal sequence: the asialoglycoprotein receptor membrane
 anchor."
 RL Cell 44:177-185 (1986).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX Wang H., Gao X., Li L., Lou H., Huang Y., Wang B., Han J.;
 RA "Human asialoglycoprotein receptor 1 gene is expressed in SH-SY5Y
 RT human neuroblastoma cells."
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Dege J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Uadin T.B., Tohiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
 RA Boeak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,

RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
 RA Schnurch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [5]
 RP INTERACTION WITH LASS2.
 RX PubMed=11543633; DOI=10.1006/geno.2001.6614;
 RA Pan H., Qin W.-X., Huo K.-K., Wan D.-P., Yu Y., Xu Z.-G., Hu Q.-D.,
 RA Gu K.-T., Zhou X.-M., Jiang H.-Q., Zhang P.-Y., Huang Y., Li Y.-Y.,
 RA Gu J.-R.;
 RT "Cloning, mapping, and characterization of a human homologue of the
 RT yeast longevity assurance gene LAG1."
 RL Genomics 77:58-64 (2001).
 CC -1- FUNCTION: Mediates the endocytosis of plasma glycoproteins to
 CC which the terminal sialic acid residue on their complex
 CC carbohydrate moieties has been removed. The receptor recognizes
 CC terminal galactose and N-acetylgalactosamine units. After ligand
 CC binding to the receptor, the resulting complex is internalized and
 CC transported to a sorting organelle, where receptor and ligand are
 CC disassociated. The receptor then returns to the cell membrane
 CC surface.
 CC -1- SUBUNIT: Interacts with LASS2.
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein.
 CC -1- TISSUE SPECIFICITY: Expressed exclusively in hepatic parenchymal
 CC cells.
 CC -1- PTM: Phosphorylated on a cytoplasmic Ser residue.
 CC -1- MISCELLANEOUS: Calcium is required for ligand binding.
 CC -1- SIMILARITY: Contains 1 C-type lectin family domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/submit/> or
 CC or send an email to license@ebi.ac.uk).
 CC -----
 CC EMBL; M10058; AA51785.1; -;
 DR EMBL; AB070933; BAB83508.1; -;
 DR EMBL; BC032130; AAH32130.1; -;
 DR PIR; A22509; LNHU1.
 DR PDB; 1DV8; X-ray; A=153-280.
 DR Genew; HGNC:742; ASGR1.
 DR H-InvDB; HIX0013487; -;
 DR MIM; 108360; -;
 DR GO; GO:0005887; C:integral to plasma membrane; TAS.
 DR GO; GO:0004873; F:asialoglycoprotein receptor activity; TAS.
 DR GO; GO:0006898; P:receptor mediated endocytosis; TAS.
 DR InterPro; IPR002353; Antifreeze1.
 DR InterPro; IPR001304; Lectin_C.
 DR Pfam; PF00059; Lectin_C; 1.
 DR Pfam; PF03954; Lectin_N; 1.
 DR PRINTS; PR00356; ANTI-FREEZE1.
 DR PROSITE; PS00615; C-TYPE-LECTIN_1; 1.
 DR PROSITE; PS00615; C-TYPE-LECTIN_2; 1.
 DR 3D-structure; Calcium; Endocytosis; Glycoprotein; Lectin;
 KW Phosphorylation; Receptor; Signal-anchor; Transmembrane.
 FT INIT MET 0 0
 FT DOMAIN 1 39 Cytoplasmic (Probable).
 FT TRANSMEM 40 60 Signal-anchor for type II membrane
 FT protein (Probable).
 FT DOMAIN 61 290 Extracellular (Probable).
 FT DOMAIN 152 278 C-type lectin.
 FT SITE 4 7 Endocytosis signal (Potential).
 FT DISULFID 153 164 By similarity.
 FT DISULFID 181 276 By similarity.

FT	DISULFID	254	268	By similarity.
FT	CARBOHYD	80	80	O-linked.
FT	CARBOHYD	148	148	O-linked.
FT	TURN	155	156	
FT	STRAND	158	160	
FT	TURN	161	162	
FT	STRAND	163	167	
FT	STRAND	172	172	
FT	HELIX	174	183	
FT	TURN	184	185	
FT	STRAND	187	188	
FT	HELIX	194	204	
FT	STRAND	209	214	
FT	TURN	216	217	
FT	STRAND	221	222	
FT	TURN	223	224	
FT	STRAND	229	230	
FT	TURN	235	235	
FT	TURN	237	238	
FT	TURN	245	246	
FT	TURN	254	257	
FT	TURN	259	260	
FT	STRAND	263	266	
FT	TURN	268	269	
FT	STRAND	272	279	
SQ	SEQUENCE	290 AA; 33055 MW; B1897CE30DAE1586 CRC64;		
Query Match				
Best Local Similarity 52.8%; Score 918.5; DB 1; Length 290;				
Matches 171; Conservative 47; Mismatches 67; Indels 29; Gaps 5;				
QY	2	TRTYENFOYLENKVK-VQGFKNGLPLQLSLLQRLRSRGPCHLLLSLGLGLLLVIVCVGF	60	
Db	1	TREYDQLHLDNEESDHDHQRKGPPLPQLQLCSGPRLLLSLGLSLLLVVVCVIGS	60	
QY	61	QNSKFFORDLVLTDPSTNTSVARIQALTSQSSLEETIASLKAIEVGFQKQERQAGS	120	
Db	61	QNSQLEELRGLRETENFTASTAQVKGSLTQGGNVGRKMSLSQLE--KQK-----	113	
QY	121	ELQEHHTQKAHLGHCHPCFSCVCPVHSEMLLRVQQLVQDLKLTQVATLNNASTEGTC	180	
Db	114	DLSED-----HSLLLHVQFVSDRLSLSCOMALQNGS-ERTC	152	
QY	181	CPVNVVEHODSCVYFSGMSMAEAEKYCOLKNAHLVWVINSREQNFQVQKYLGSAYTWNG	240	
Db	153	CPVNVVEHRSYCFYFSRSGKAWADADNYCRLEDAHLVVVTSWEEQKFVQHHIGPVNTWNG	212	
QY	241	LSDPEGANKVVDGTYATQFQWKQPDQWQHGGLGGEDCAHFHPDGRWDDVCORPY	300	
Db	213	LHDQNGPFWKVDGTYETGFKWVRPQPDQWYGHGLGGEDCAHFHTDDGRWDDVCORPY	272	
QY	301	HWVCEAGLGQTSQE	314	
Db	273	RWCETELDKASQE	286	
RESULT 5				
MMGL_RAT				
ID	MMGL_RAT	STANDARD;	PRT;	306 AA.
AC	P49301;			
DT	01-FEB-1996 (Rel. 33, Created)			
DT	01-FEB-1996 (Rel. 33, Last sequence update)			
DT	05-JUL-2004 (Rel. 44, Last annotation update)			
DE	Macrophage asialoglycoprotein-binding protein (M-ASGP-Bp) (Macrophage			
DE	Galactose/N-acetylgalactosamine-specific lectin) (MMGL).			
GN	Name=Mg11; Synonyms=Mg1;			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_TaxID=10116;			
RN	[1]			
RP	SEQUENCE FROM N.A.; AND PARTIAL SEQUENCE.			
RX	MEDLINE=90293078; PubMed=2358462;			
RA	Ii M., Kurata H., Itoh N., Yamashina I., Kawasaki T.;			
RT	"Molecular cloning and sequence analysis of cDNA encoding the			
RT	macrophage lectin specific for galactose and N-acetylgalactosamine.;"			
RL	J. Biol. Chem. 265:11295-11298(1990).			
RL	[2]			
RN	PRELIMINARY SEQUENCE OF 9-28.			
RA	MEDLINE=88339956; PubMed=3421964;			
RX	Ii M., Kawasaki T., Yamashina I.;			
RT	"Structural similarity between the macrophage lectin specific for			
RT	galactose/N-acetylgalactosamine and the hepatic asialoglycoprotein			
RT	binding protein.;"			
RL	Biochem. Biophys. Res. Commun. 155:720-725(1988).			
CC	-!- FUNCTION: Recognizes terminal galactose and N-acetylgalactosamine			
CC	units.			
CC	-!- SUBUNIT: Homooligomer.			
CC	-!- SUBCELLULAR LOCATION: Type II membrane protein.			
CC	-!- SIMILARITY: Contains 1 C-type lectin family domain.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; J05495; AAA41216.1; -.			
DR	PIR; A42230; A42230.			
DR	HSP; P07306; IDV8.			
DR	InterPro; IPR002353; AntifreezeII.			
DR	InterPro; IPR001304; Lectin_C.			
DR	InterPro; IPR005640; lectin_N.			
DR	InterPro; IPR009053; Prefoldin.			
DR	Pfam; PF00059; Lectin_C; 1.			
DR	Pfam; PF03954; Lectin_N; 1.			
DR	PRINTS; PR00356; ANTIFREEZEII.			
DR	SMART; SM00034; CLECT; 1.			
DR	PROSITE; PS00615; C-TYPE LECTIN_1; 1.			
DR	PROSITE; PS00641; C-TYPE LECTIN_2; 1.			
KW	Calcium; Direct protein sequencing; Glycoprotein; Lectin;			
KW	Signal-anchor; Transmembrane.			
FT	DOMAIN 1 37			
FT	TRANSMEM 38 58			
FT	DOMAIN 59 306			
FT	DOMAIN 174 300			
FT	DISULFID 175 186			
FT	DISULFID 203 298			
FT	DISULFID 276 290			
FT	CARBOHYD 76 76			
FT	CARBOHYD 168 168			
SQ	SEQUENCE 306 AA; 34242 MW; D68A5DFF0B9B8F13 CRC64;			
Query Match				
Best Local Similarity 51.0%; Score 887; DB 1; Length 306;				
Matches 177; Conservative 38; Mismatches 80; Indels 28; Gaps 5;				
QY	1	MTRTYENFOYLENKVKVQGFKNGLPLQLSLLQRLRSRGPCHLLLSLGLGLLLVIVCVGF	60	
Db	1	MTMAYENFQNLGSEKQ--EAGKAPQPSFLCNLSWTHLLLSLGLSLLLVISVIGS	58	
QY	61	QNSKFFORDLVLTDPSTNTSVARIQALTSQSSLEETIASLKAIEVGFQKQERQAG--	118	
Db	59	QNSQLRRDLTLRTLDNTTNTKALQALASRGDSLQTLGINSLKVEVDHGHQELQAGRG	118	
QY	119	-----VSEIQ-----EHTTQKAHLGHCHPCFSCVCPVHSEMLLRVQQLVQDLKLTQVAT	169	
Db	119	LSQKVASLESTVEKEQLRTDL-----SEITDRVQQLGKDLKLTLCQLAS	164	
QY	170	LNNASTEGTCPPVNVVEHODSCVYFSGMSMAEAEKYCOLKNAHLVWVINSREONFVO	229	
Db	165	LKNNGSAV-ACCPHLHMEHSGCYWFSGKFPWPEADKYCOLENSLVVVNSLAEQNFQ	223	

```

QY 230 KYLGSAITWGLSDPEGAWKWVDGTYATGCFQNKPCQDDWQGHGLGGEDCAHFHPDG 289
Db 224 THMGSVVTWGLTDQNGPRWVDGTYEKGFTTHWAPQPDNMYGHGLGGEDCAHFSTDG 283
QY 290 RWNDDVCQRPYHWCBAAGLQOTS 312
Db 284 RWNDDVCQRPYHWCBAAGLQOTS 306

RESULT 6
QY1Y84
ID QY1Y84 PRELIMINARY; PRT; 284 AA.
AC QY1Y84;
DT 01-DEC-2001 (T-EMBLrel. 19, Created)
DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)
DT 25-OCT-2004 (T-EMBLrel. 28, Last annotation update)
DE Asialoglycoprotein receptor major subunit (Asialoglycoprotein receptor
DE 1).
GN Name=ASG1;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SvJ; TISSUE=Liver;
RX MEDLINE=20137499; PubMed=10675034; DOI=10.1016/S0378-1119(99)00493-X;
RA Soukharev S., Berlin W., Hanover J.A., Bethke B., Sauer B.;
RT "Organization of the mouse ASG1 gene encoding the major subunit of
RT the hepatic asialoglycoprotein receptor.";
RL Gene 241:233-240(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Liver;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan T., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettaman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Liver;
RA Strausberg R.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF182811; AAF29495.1; -
DR EMBL; BC022106; AAH22106.1; -
DR PIR; S29855; S29855.
DR HSSP; P07306; IDV8.
DR MGD; MGI-88081; Asg1.
DR GO; GO:0016021; C: integral to membrane; TAS.
DR Pfam; PF00059; Lectin_C; 1.
DR Pfam; PF03954; Lectin_N; 1.
DR PRINTS; PR00356; ANTIREEZEII.
DR SMART; SM00034; CLECT; 1.
DR PROSITE; PS00615; C_TYPE_LLECTIN_1; 1.
DR PROSITE; PSS0041; C_TYPE_LLECTIN_2; 1.
KW Receptor.

```

```

SQ SEQUENCE 284 AA; 32591 MW; 671F043968047DB7 CRC64;
Query Match 50.5%; Score 877; DB 2; Length 284;
Best Local Similarity 51.3%; Pred. No. 4.6e-65;
Matches 160; Conservative 54; Mismatches 70; Indels 28; Gaps 4;

QY 1 MTRTYENFOYLENKVKVQGFQKNGPLQSLQRLQRLSGPCCHLLSLGLGLLLVIVCVGF 60
Db 1 MYKDYQDFHLDNDNDHQLRRGPPPTPRLLQLKCSGSRLLSSLSILLVIVCVITS 60
QY 61 QNSKFORDLVTLRTDFSNFTSNTVAEIQALTSQSSLEETIASLKAEVEGFQKQAGVS 120-
Db 61 QNSQLREDLLALRQNFENLTVSTEDQVKALSTQSSVGRKMKLVESKLE--KQK----- 113
QY 121 ELQETTTQXAHGLGCHPCFSVCVPVHSEMLLRVQVLQDLKKLTQCVATLNNASTEGTC 180
Db 114 DLTED-----HSLLLHVKQLVSDVRSLSQMAAFRNGS-ERTC 152
QY 181 CPWNVVEHODSCYFWSHSGMSMAEAKYCOLKNAHLVWINSREQNFVOKYLGSAVTWG 240
Db 153 CPINWVEYEGSCYFWSVSRPWTEDAKYQLENAHLVWVTSRDEQFLQRMGLNTWIG 212
QY 241 LSDPEGAWKWVDGTYATGCFQNKPCQDDWQGHGLGGEDCAHFHPDGRWDDVCORPY 300
Db 213 LTDQNGPKWVDGTYETGCFQNWPRPQPDNMYGHGLGGEDCAHFSTDGRRNDVCRPY 272
QY 301 HWCEAGLGQTS 312
Db 273 RWVCEKLDKAN 284

RESULT 7
ID LECH RAT STANDARD; PRT; 283 AA.
AC P02706;
DT 21-JUL-1986 (Rel. 01, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Asialoglycoprotein receptor 1 (Hepatic lectin 1) (RHL-1) (ASGP-R)
DE (ASGPR).
GN Name=ASG1; Synonyms=Asgr-1;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86008335; PubMed=2995379;
RA Leung J.O., Holland E.C., Drickamer K.;
RT "Characterization of the gene encoding the major rat liver
RT asialoglycoprotein receptor.";
RL J. Biol. Chem. 260:12523-12527(1985).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=85063786; PubMed=6095287;
RA Holland E.C., Leung J.O., Drickamer K.;
RT "Rat liver asialoglycoprotein receptor lacks a cleavable NH2-terminal
RT signal sequence.";
RL Proc. Natl. Acad. Sci. U.S.A. 81:7338-7342(1984).
RN [3]
RP SEQUENCE OF 11-283 FROM N.A.
RX MEDLINE=87026895; PubMed=2945599;
RA Watts C.;
RT "Isolation and expression of cDNA clones for a rat liver
RT asialoglycoprotein receptor.";
RL Biosci. Rep. 6:527-534(1986).
CC -1- FUNCTION: Mediates the endocytosis of plasma glycoproteins to
CC which the terminal sialic acid residue on their complex
CC carbohydrate moieties has been removed. The receptor recognizes
CC terminal galactose and N-acetylgalactosamine units. After ligand
CC binding to the receptor, the resulting complex is internalized and
CC transported to a sorting organelle, where receptor and ligand are
CC disassociated. The receptor then returns to the cell membrane

```


FT	TRANSMEM	39	59	Signal-anchor for type II membrane protein (Potential).
FT	DOMAIN	60	283	Extracellular (Potential).
FT	DOMAIN	151	277	C-type lectin.
FT	SITE	4	7	Endocytosis signal (Potential).
FT	DISULFID	152	163	By similarity.
FT	DISULFID	180	275	By similarity.
FT	DISULFID	253	267	By similarity.
FT	CARBOHYD	74	77	N-linked (GLCNAC. . .) (Potential).
FT	CARBOHYD	77	77	N-linked (GLCNAC. . .) (Potential).
FT	CARBOHYD	145	145	N-linked (GLCNAC. . .) (Potential).
FT	CONFLICT	150	150	I -> T (in Ref.1).
FT	SEQUENCE	283 AA;	32472 MW;	982A5D305AAE0D8F CRC64;
Query Match		49.8%;	Score 866;	DB 1; Length 283;
Best Local Similarity		50.8%;	Pred. No. 3.8e-64;	
Matches 158;	Conservative	54;	Mismatches 71;	Indels 28; Gaps 4;
QY	2	TRTYENFOYLENKVKVQGFNGPLPQSLQRLRSRGPCHILLSLGLGLLLVITCVGVGF	61	
Db	1	TKDYQFOHLDNDNDHQLRRRGGPPPTRLQLRCGSRLLLSLLSSILLVWVVCVITSQ	60	
QY	62	NSKFORDLVLRDRFSNTVAETQALTSQSSLEETIASLKAAYEGFKQBRQAGVSE	121	
Db	61	NSQLREDLALRNFENLTVSTEDQVKALSTQSSVGRKWKLVESKUE--KQQK-----D	113	
QY	122	LOEHHTTQKAHLGCHPCPSVCPVHSEMLLRVQQLVQDLKKLTQCVATLNNASTEGTCC	181	
Db	114	LTED-----HSSLHLHVQLVSDVRSLSQMAAFRNGS-ERICC	152	
QY	182	PVNVEHQDSVWFSGHSGMSWAEEKYCOLKNAHLVVINSREDFONFVKYLGSAYTWMLG	241	
Db	153	PINWVEYEGSCVWFSSVRPWTADKYCYOLENAHLVVVTSRDEQNFQLQRHMGPLNTWIGL	212	
QY	242	SDPEGAKWVDGTDYATGQNWKPQGDQHQHGLGGEDCAHPHDPGRWNDDVCORVPH	301	
Db	213	TDQNGPWWKVDGTDYETGFQNWPRPEQPDNNWYHGLGGEDCAHFTTDGRMNDVCRRPYR	272	
QY	302	WVCEAGLAGQTS	312	
Db	273	WVCETKLDKAN	283	
RESULT 10				
Q91Y73				
ID	Q91Y73	PRELIMINARY;	PRT;	304 AA.
AC	Q91Y73;			
DT	01-DEC-2001	(TEMBLrel. 19, Created)		
DT	01-DEC-2001	(TEMBLrel. 19, Last sequence update)		
DT	01-MAR-2004	(TEMBLrel. 26, Last annotation update)		
DE	Macrophage galactose N-acetyl-galactosamine specific lectin 1.			
GN	Name:Mgl1;			
OS	Mus musculus (Mouse)			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=FVB/N; TISSUE=Mammary tumor;			
RC	MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;			
RX	Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,			
RA	Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,			
RA	Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,			
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh F.,			
RA	Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,			
RA	Stapleton M., Soares M.B., Bonaldo M.F., Casavint T.L., Scheetz T.E.,			
RA	Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,			
RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,			
RA	Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,			
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,			
RA	Villalón D.K., Muzy D.M., Sodergren E.J., Lu X., Gibbs R.A.,			
RA	Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,			
RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,			

Db 297 GRWNDDVCRHWHVICETELGKAS 320

RESULT 9
LECH MOUSE STANDARD; PRT; 283 AA.
ID LECH_MOUSE STANDARD; PRT; 283 AA.
AC P34927; Q64363;
DT 01-FEB-1994 (Rel. 28, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Asialoglycoprotein receptor 1 (Hepatic lectin 1) (MHL-1) (ASGP-R) (ASGPR).
DE Name=Asgr1; Synonyms=Asgr-1;
GN Mus musculus (Mouse).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
[1]
RN SEQUENCE FROM N.A.
RP MEDLINE=93176818; PubMed=8439566; DOI=10.1016/0167-4781(93)90300-3;
RX Takezawa R., Shinzawa K., Watanabe Y., Akaike T.;
RA "Determination of mouse major asialoglycoprotein receptor cDNA
sequence.";
RT Biochim. Biophys. Acta 1172:220-222(1993).
RL [2]
RN SEQUENCE FROM N.A.
RP STRAIN=BALB/c; TISSUE=Liver;
RX MEDLINE=95047431; PubMed=7958950; DOI=10.1016/0378-1119(94)90694-7;
RA Monroe R.S., Huber B.E.;
RT "The major form of the murine asialoglycoprotein receptor: cDNA
sequence and expression in liver, testis and epididymis.";
RL Gene 148:237-244(1994).
CC -!- FUNCTION: Mediates the endocytosis of plasma glycoproteins to
which the terminal sialic acid residue on their complex
carbohydrate moieties has been removed. The receptor recognizes
terminal galactose and N-acetylgalactosamine units. After ligand
binding to the receptor, the resulting complex is internalized and
transported to a sorting organelle, where receptor and ligand are
disassociated. The receptor then returns to the cell membrane
surface.
CC -!- SUBUNIT: Interacts with LASS2 (By similarity).
CC -!- SUBCELLULAR LOCATION: Type II membrane protein.
CC -!- TISSUE SPECIFICITY: Expressed exclusively in hepatic parenchymal
cells.
CC -!- MISCELLANEOUS: Calcium is required for ligand binding.
CC -!- SIMILARITY: Contains 1 C-type lectin family domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to licensese@isb-sib.ch).

DR EMBL; D13517; BAA02734.1; -;
DR EMBL; U09362; AAB60441.1; -;
DR EMBL; U08372; AAB60440.1; -;
DR HSSP; P07306; LDV8.
DR MGD; MGI:88081; Asgr1.
DR InterPro; IPR002353; Antifreezezf.
DR InterPro; IPR001304; Lectin_C.
DR InterPro; IPR005640; Lectin_N.
DR Pfam; PF000059; Lectin_C_1.
DR Pfam; PF03954; Lectin_N_1.
DR PRINTS; PR00356; ANTIFREEZEII.
DR SMART; SM00034; CLECT; 1.
DR PROSITE; PS00615; C TYPE LECTIN 1; 1.
DR PROSITE; PS50041; C TYPE LECTIN 2; 1.
KW Calcium; Endocytosis; Glycoprotein; Lectin; Phosphorylation; Receptor;
Signal-anchor; Transmembrane.
KW INIT MET 0 0 By similarity.
FT DOMAIN 1 38 Cytoplasmic (Potential).

FT TRANSMEM 39 59 Signal-anchor for type II membrane
FT FT protein (Potential).
FT DOMAIN 60 283 Extracellular (Potential).
FT DOMAIN 151 277 C-type lectin.
FT SITE 4 7 Endocytosis signal (Potential).
FT DISULFID 152 163 By similarity.
FT DISULFID 180 275 By similarity.
FT CARBOHYD 74 74 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 77 77 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 145 145 N-linked (GlcNAc...) (Potential).
FT CONFLICT 150 150 I -> T (in Ref. 1).
SQ SEQUENCE 283 AA; 32472 MW; 982ASD30SAE0D8F CRC64;

Query Match 49.8%; Score 866; DB 1; Length 283;
Best Local Similarity 50.8%; Pred. No. 3.8e-64;
Matches 158; Conservative 54; Mismatches 71; Indels 28; Gaps 4;

QY 2 TRTYENFQYLENKVKVQGFNGPLFLOSLQLRSGPCHELLLSGLGLLVITCVVGFG 61
DB 1 TKDYQDFOLNDNDHHQLRRGGPTPLRLQLRCGSRLLLSLSILLVVVCVITSQ 60
QY 62 NSKFQFDLYLTRDFSNTVAEQIALTSOGSSLEETIASLKAEVGFQERQAGVSE 121
DB 61 NSOLREDLLALQNFSNLTVSTEDQVKALSTUGSSVGRMKLVESKLJB--KKQK----D 113
QY 122 LOEHHTTKAHLGHCPFCPSVCPVHSEMRLRVQQLVODLKKLTTCOVATLNNNASTEGTCC 181
DB 114 LTED-----HSSLHLHVQLVSDVRSLSQCMAPFRNGS-ERICC 152
QY 182 PVNVVEHQSCYWFHSHGSMWAEASKYCOLKNVHLVINSRBEQNFVKYLGSAVTWMGL 241
DB 153 PINWVEYGSCYWFSSVRPWTEADKYCOLENAHLVVVTSRDEQNFLOHRMGPLNWLGL 212
QY 242 SPEGGAWKWDGTDYATGFWKPGOPDDWGCGHGGGEDCAHFHPDWNDVDCORPVH 301
DB 213 TTQONGPKWVDGTDYETGFWNRPEQPNWYHGLGGGEDCAHFTTGTGWRNDDVCCRPR 272
QY 302 WVCEAGLGQTS 312
DB 273 WVCETKLDKAN 283

RESULT 10
QY1YT3 PRELIMINARY; PRT; 304 AA.
ID QY1YT3
AC QY1YT3;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE Macrophage galactose N-acetyl-galactosamine specific lectin 1.
GN Name=Mgl1;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=FVB/N; TISSUE=Mammary tumor;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Munzy D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,


```

QY 179 TCCPVNVVHODSCYWFSGHSMWAEKCYCOLKNAHLVINSREBQNFVOKYLGSAITW 238
DB 171 ACCPLHTEHEGSCYWFSEKSWPADKYCRLENSHLVVSLEBQNFQNLANVSW 230
QY 239 MGLSDPEGAKWVDGTDYATGFQWPKPGQDDQGHGLGGEDCAHFHPDGRWDDVCQR 298
DB 231 IGLTDQNGPWRWDGTDGDFEKGFKFNAPLQPDNFGHGLGGEDCAHITGGPWNDDVCQR 290
QY 299 PYHWVCEAGLGQTS 312
DB 291 TFRWICEMKLAKES 304

RESULT 12
Q8BUD5 PRELIMINARY; PRT; 361 AA.
AC Q8BUD5;
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
DE Mus musculus 10 days lactation, adult female mammary gland cDNA, RIKEN
DE full-length enriched library, clone:D730047H02 product:similar to
DE MACROPHAGE GALACTOSE N-ACETYL-GALACTOSAMINE SPECIFIC LECTIN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Mammary gland;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
[2]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Mammary gland;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
[3]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Mammary gland;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Akiyama J., Nishi K., Kiteunai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
[5]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Mammary gland;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoaka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kiteunai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).

```

```

RN [6]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Mammary gland;
RA Adachi J., Aizawa K., Akahira S., Fukuda S., Fukunishi Y., Furuno M.,
RA Arakawa T., Bono H., Carninci P., Hiramoto K., Hiraoka T., Hori F.,
RA Hanagaki T., Hara A., Hayatsu N., Izawa M., Kasukawa T., Kato H.,
RA Imotani K., Ishii Y., Itoh M., Kouda M., Koya S., Kurihara C.,
RA Kawai J., Kojima Y., Konno H., Kouda M., Kouda M., Numazaki R., Ohno M.,
RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,
RA Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,
RA Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,
RA Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
RA Tejima Y., Toya T., Yamamura T., Yamanaka I., Yasunishi A.,
RA Yoshida K., Yoshino M., Muramatsu M., Hayashizaki Y.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK085751; BAC39530.1; -.
DR HSSP; P07306; 1DV8.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005529; F:sugar binding; IEA.
DR InterPro; IPR001304; LECTIN_C.
DR InterPro; IPR005640; LECTIN_N.
DR InterPro; IPR009053; Prefoldin.
DR Pfam; PF00059; LECTIN_C; 1.
DR Pfam; PF03954; LECTIN_N; 1.
DR SMART; SM00034; CLECT; 1.
DR PROSITE; PS00615; C_TYPE_LLECTIN_1; 1.
DR PROSITE; PS0041; C_TYPE_LLECTIN_2; 1.
KW LECTIN.
SQ SEQUENCE 361 AA; 41010 MW; AACBBY48C66A231 CRC64;

Query Match 45.5%; Score 791; DB 2; Length 361;
Best Local Similarity 47.0%; Pred. No. 9.8e-58;
Matches 158; Conservative 40; Mismatches 86; Indels 52; Gaps 3;

QY 25 LPLQSLQLRSGPCCHLLSLGLGLLLVLCVVGQNSKQFQDLVTLRTDFSNFTSNTV 84
DB 18 IPSQSLWRLSWTHLLFSLGLSLLLVVISVIGSQNSQLRDLGLTILAILDNTTSKIK 77
QY 85 AEIQALTSQSSLEETIASLKAEVGFQKQERQAGVSELOHTQKAHLGHCHPCSPVCP 144
DB 78 AEFQSLDSRADNFEKGISSLVKVDVEDHRLQAGRLDSQKVTSLSTLEKREKALKTDL- 136
QY 145 VHSEMLLRVQQLVODLKLKLTQVATLNNASTGTCTCPVNVVHODSCYWFSGHSMWAE 204
DB 137 --SDLTDHVVQQLTDLKALTQCLANLKNNGS-EVACCPHLWTEHEGSCYWFSEKSWPE 193
QY 205 AEKYCOLKNAHLVINSREBQNFVOKYLGSAITWGLSDPEGAKWVDGTDYATG----- 259
DB 194 ADKYCELENSHLVVSLEBQNFQNLANVLSWMLTDQNGPWRWDGTDGDKGPKYVC 253
QY 260 -----FQNKPGQDDQGHGL 276
DB 254 RLQLAPLYLGLSYLSPDFSDPDLGPGSGNMADGQIWSAQOFFIFRNWRPLQPDNWHGML 313
QY 277 GGGEDCAHFHPDGRWDDVCQRYHWVCEAGLGQTS 312
DB 314 GGGEDCAHFSDYGRWDDVCQRYHWVCEAGLGQTS 349

RESULT 13
Q7Z4G9 PRELIMINARY; PRT; 306 AA.
AC Q7Z4G9;
DT 01-OCT-2003 (T-EMBLrel. 25, Created)
DT 01-OCT-2003 (T-EMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
DE HbxAg-binding protein.
GN Name=HBXBP;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]

```


C

Db 213 GLHDQNGPFWKWDGTDYETGKWRPEQPDWYGHGLGGEDCAHFTDDGRWNDDVCORP 272
QY 300 YHWVCEAGLGQTSQE 314
Db 273 YRWVCETELDKASQE 287

RESULT 2
A42230
lectin_M-ASGP-BP precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 19-Jun-1992 #sequence_revision 19-Jun-1992 #text_change 09-Jul-2004
C:Accession: A42230
R;14, M.; Kurata, H.; Itoh, N.; Yamashina, I.; Kawasaki, T.
J. Biol. Chem. 265, 11295-11298, 1990
A:Title: Molecular cloning and sequence analysis of cDNA encoding the macrophage lectin
A:Reference number: A42230; MUID:90293078; PMID:2358462
A:Accession: A42230
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-306 <11A>
A:Cross-references: UNIPROT:P49301; GB:J05495; NID:G204302; PIDN:AAA41216.1; PID:G204303
C:Superfamily: hepatic lectin; C-type lectin homology
F;175-298/Domain: C-type lectin homology <LCH>

Query Match 51.0%; Score 887; DB 2; Length 306;
Best Local Similarity 54.8%; Pred. No. 1.5e-64;
Matches 177; Conservative 38; Mismatches 80; Indels 28; Gaps 5;

QY 1 MTRTYENFOYLENKVKVQGFKNQGLPLQSLQRLQRLSRGCHLLSLGLGLGLLLVILCVVGF 60
Db 1 MTWAYENFONLGSSEKQ--EAGKAPQSFCLNLTSLTHLLFSLGLSLLLVVISVIGS 58
QY 61 QNSKFQDRLVLRDTSFNFTSNVVAETQALTSQSSLEETIASLKARVEGFKQERQAG-- 118
Db 59 QNSQLRDLLETRTLTLDNTTNTKAELQALASRGDSLQGTGINSIKVEVDVHGQELQAGRG 118
QY 119 ----VSLEQ-----EHTTKAHLGHCPHCPSCVCPVPHSEMLLRVQQLVQDLKLTQVAT 169
Db 119 LSQKVASLESTVEKEQTLRTDL-----SEITDRVQKGLKLTQLAS 164
QY 170 LNNNASTEGTCCPVNMYEHQDSCYWFHSGMSWAEAEKYCOLKNAHLVWINSREEQNFKV 229
Db 165 LKNGSAV-ACCPHLHWEHSGCYWFSQSKPWEADKYQLENSNLVNVNSLAEQNFLQ 223
QY 230 KYLGSAVTWNGLSDPGAWKWDGTDYATGFQNKWKPQDDWQGHGLGGEDCAHFTPDG 289
Db 224 THMGSVVTVTGLTDQNGFWRVWDGTDYKGFTHWAPKPDNWTYGHGLGGEDCAHFTSDG 283
QY 290 RWNDVVCORPVHWVCEAGLGQTS 312
Db 284 RWNDVVCORPYRWVCEMKLAKDS 306

RESULT 3
LNRTL
hepatic lectin - rat
N:Alternate names: ASGP; asialoglycoprotein receptor
C:Species: Rattus norvegicus (Norway rat)
C:Date: 28-Feb-1986 #sequence_revision 04-Dec-1986 #text_change 09-Jul-2004
C:Accession: A92497; A94020; B94020; A54727; A03166
R;Leung, J.O.; Holland, E.C.; Drickamer, K.
J. Biol. Chem. 260, 12523-12527, 1985
A:Title: Characterization of the gene encoding the major rat liver asialoglycoprotein re
A:Reference number: A92497; MUID:86008335; PMID:2995379
A:Accession: A92497
A:Molecule type: DNA
A:Residues: 1-284 <LEU>
A:Cross-references: UNIPROT:P02706; GB:K02817; NID:G206646; PIDN:AAA42037.1; PID:G206647
R;Holland, E.C.; Leung, J.O.; Drickamer, K.
Proc. Natl. Acad. Sci. U.S.A. 81, 7338-7342, 1984
A:Title: Rat liver asialoglycoprotein receptor lacks a cleavable NH-2-terminal signal se

Db 61 LRRDLGTLRLATLDNTTISKIAEFQSLDRSDRAISFEKGISLSLKVDVDEHRELQAGRDLQSK 120

Qy 119 VSELOEHTTKAKHLGHCPHCPSVCVPVHSEMLLRVQQLVQDILKLTCQVATLNNASTEG 178
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
121 VTSL-ESTVEKRE-----QALKTDLSLDTDHVQQLRKDKLKALTCQLANLNKNNGS-EV 170

Qy 179 TCCPWNVVEHQDSCTWFSHSGMSWAEEAKYICOLKNAHLVVINSRBEONFVKYLGSAYTW 238
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
171 ACCPLAHTHEHSGCYWFSESSEKSPKADKYCRLENSHLVVVNSSLSEQNFLQRLANVSW 230

Qy 239 MGLSDPEGAWKWDGTGYATCFQNWKKPCQPDPDWOGHGLGGGDCAHFHPDGGRWDDVCCR 298
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
231 IGLTDONGPRWVDGTDFEKGFKNAPLPDNWFGHGLGGGEDCAHIITCGGPWDDVCCR 290

Qy 299 PYHWVCBAGLGQTS 312
: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 291 TFRWICEMKLAKES 304

RESULT 6
LNHUZA
asialoglycoprotein receptor H2a - human
N/Alternate names: hepatic lectin H2a
N/Contains: ASGPR; asialoglycoprotein receptor H2a; asialoglycoprotein receptor H2b
C/Species: Homo sapiens (man)
C/Date: 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 09-Jul-2004
C/Accession: A25179; B39100; B39100; I37995; A49466; B49466; S14525
R/Spies, M.; Lodish, H.F.
Proc. Natl. Acad. Sci. U.S.A. 82, 6465-6469, 1985
A/Title: Sequence of a second human asialoglycoprotein receptor: conservation of two rec
A/Reference number: A25179; MUID:86016723; PMID:3863106
A/Acession: A25179
A/Molecule type: mRNA
A/Residues: 1-311 <SPI>
A/Cross-references: UNIPROT:P07307; GB:M11025; NID:g179080; PIDN:AAB59519.1; PID:g179081
R/Lederkremer, G.Z.; Lodish, H.F.
J. Biol. Chem. 265, 1237-1244, 1991
A/Title: An alternatively spliced minixon alters the subcellular fate of the human asia
A/Reference number: A39100; MUID:91093236; PMID:1985943
A/Acession: A39100
A/Molecule type: DNA; mRNA
A/Residues: 69-99 <LED>
A/Cross-references: GB:M38420; NID:g184395
A/Acession: B39100
A/Molecule type: DNA; mRNA
A/Residues: 69-81,87-99 <LE2>
A/Cross-references: GB:M38420; NID:g184395
R/Paletta, E.; Stockert, R.J.; Racevskis, J.
Hepatology 15, 395-402, 1992
A/Title: Differences in the abundance of variably spliced transcripts for the second asi
A/Reference number: I37995; MUID:92184202; PMID:I371982
A/Acession: I37995
A/Status: preliminary; translated from GB/EMBL/DDBJ
A/Molecule type: mRNA
A/Residues: 1-23,43-81,87-311 <PAI>
A/Cross-references: EMBL:X55283; NID:G34354; PIDN:CAA38997.1; PID:G34355
R/Yuk, M.H.; Lodish, H.F.
J. Cell Biol. 123, 1735-1749, 1993
A/Title: Two pathways for the degradation of the H2 subunit of the asialoglycoprotein rec
A/Reference number: A49466; MUID:94103329; PMID:8276894
A/Acession: A49466
A/Molecule type: protein
A/Residues: 78-98 <YUK>
A/Acession: B49466
A/Molecule type: protein
A/Residues: 87-98 <YU2>
C/Comment: The functioning ligand-binding unit of this receptor is thought to be at least
C/Genetics:
A/Gene: GDB:ASGR2; L-H2
A/Cross-references: GDB:118755; OMIM:108361
A/Map position: 17p13-17p11
C/Superfamily: hepatic lectin; C-type lectin homology

QY 162 KLTQVATLNNASTETGCPVNVVHEDQSCYFWSHSGMSWAEAKYCOLKNAHLVINS 221
Db 154 TLTCQLAFLFNGT---ECCPVNVVFGGSCYFSDRLGTWAEADQYQCMENAHLLVINS 210
QY 222 REBQNFQVKYLGSAITWMLSDPEGAWKWDGTDYATGFQNFQWKPQDQWQHGGLGGED 281
Db 211 REBQEFVVKHGAFAHIWGLTKDGSWKWDGTEYRSNFKNWAFTQPDNQWQHEEGGED 270
QY 282 CAHFHPDGRNDVQCRPYHWVCE 305
Db 271 CABILSDGLNDNFQCOQVNRWACE 294

RESULT 9
A28166
Kupffer cell receptor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 30-Sep-1989 #sequence_revision 30-Sep-1989 #text_change 09-Jul-2004
C;Accession: A38674; A28166
R;Hoyle, G.W.; Hill, R.L.
J. Biol. Chem. 286, 1850-1857, 1991
A;Title: Structure of the gene for a carbohydrate-binding receptor unique to rat Kupffer
A;Reference number: A38674; MUID:91107689; PMID:1846367
A;Accession: A38674
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-550 <HO2>
A;Cross-references: UNIPROT:P10716; GB:M55532; NID:g203362; PIDN:AAA40892.1; PID:g203363
R;Hoyle, G.W.; Hill, R.L.
J. Biol. Chem. 263, 7487-7492, 1988
A;Title: Molecular cloning and sequencing of a cDNA for a carbohydrate binding receptor
A;Reference number: A28166; MUID:88227939; PMID:2836387
A;Accession: A28166
A;Molecule type: mRNA
A;Residues: 1-550 <HOY>
A;Cross-references: GB:J03734; NID:g205050; PIDN:AAA41472.1; PID:g205051
A;Keywords: transmembrane protein
C;Keywords: transmembrane protein
F;412-536/Domain: C-type lectin homology <LCH>

Query Match 18.8%; Score 326.5; DB 2; Length 550;
Best Local Similarity 31.1%; Pred. No. 1.1e-18;
Matches 79; Conservative 32; Mismatches 94; Indels 49; Gaps 8;

QY 62 NSKFP---RDLVLTDFSNFTNTVAIOALTSQSSLEETIASLKAEVGFQKQRQAG 118
Db 323 NGKLDSSRELQTLRLDLS-----DVSALKSNVQMLQSLQKAKAEVQSLKTGLEA- 373
QY 119 VSELQHTTQKHLGCHPCPSVCVPVHSEMLRVQQLVODLKLITCQVATLNNASTEG 178
Db 374 -----TKTLAAKIQQQSDLEALOKAVAAHTQGOXTON 406
QY 179 TCCPV---NWVEHQDSCYFWSHSGMSWAEAKYCOLKNAHLVINSREEQNFVQKYLGSA 235
Db 407 QVLQLIMQDWKYNKGFYFSRDKSKWHEAENFCVSGAHLASVTSQEEQAFVLQITNAV 466
QY 236 YTWMLGSD--PEGAWKWDGT--DYATGFQNFQWKPQDQWQHGGLGGEDCAHFHPDGRW 291
Db 467 DHWIGLTDQCTEGNWRWVDGTPEDYVQSRFRFKGQPDNWR-HGNGEREDCVHL--QRMW 523
QY 292 NDDVCQCRPYHWVCE 305
Db 524 NDMACTAYNVNCK 537

RESULT 10
A46274
HIV gp120-binding C-type lectin - human
C;Species: Homo sapiens (man)
C;Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
C;Accession: A46274
R;Curtis, B.M.; Scharnowske, S.; Watson, A.J.
Proc. Natl. Acad. Sci. U.S.A. 89, 8356-8360, 1992
A;Title: Sequence and expression of a membrane-associated C-type lectin that exhibits CD

A;Reference number: A46274; MUID:92390446; PMID:1518869
A;Accession: A46274
A;Status: preliminary
A;Molecule type: nucleic acid
A;Residues: 1-404 <CUR>
A;Cross-references: UNIPROT:Q9NNX6
A;Experimental source: placenta
A;Note: sequence extracted from NCBI backbone (NCBIN:113134, NCBI:113135)
F;256-377/Domain: C-type lectin homology <LCH>

Query Match 18.2%; Score 316; DB 2; Length 404;
Best Local Similarity 27.1%; Pred. No. 5.3e-18;
Matches 105; Conservative 52; Mismatches 116; Indels 114; Gaps 18;

QY 16 KVQGFK-----NGPLPQSLQLRLRSGPCHLLLSLGL--GULLLVICVGVGFQNSKFPQ 66
Db 27 QTRGYKSLAGCLGHGPLVLQ-----LLSFTLLAGLLVQVSKVPSSISQFSQR 73
QY 67 RDLVLTDFSNFTNTVA-----EIQALTSQSSLEETIASL--KAEVGFQKQ-- 114
Db 74 QDAI-----YONLTQKAAVGESEKSLQEIYQELTQKAAVGEPEKSKLQEIYQELT 128
QY 115 -RQAGVSEL-----QEHTTQKHLGCHPCPSVCVPVHSEM----- 149
Db 129 RLKAAVGEPEKSKLQEIYQELTQKAAVGEPE-KSKQEIYQELTRLKAAVGEPEKS 187
QY 150 -----LLRV-----QOLVODLKKLTCQVATLNN-----NA 174
Db 188 KQEIYQELTRLKAAVGEPEKSKQEIYQELTRLKAAVGEPEKSKQEIYQELTQK 247
QY 175 STGTGTC--CPVNVVHEDQSCYFWSHSGMSWAEAKYCOLKNAHLVINSREEQNF--VOK 230
Db 248 AVERLCHPCPWEWTFPGGNCYFMSNQRNWDHSITACKEVGAQLVWIKSAEONFLQLS 307
QY 231 YLGSATVWMLGSD--PEGAWKWDGTDYATGF--QNNKPGOPDDWQHGGLGGEDCAHFHP 287
Db 308 SRNRTFWMLGSLDNLQEGTWQVDSPLPSFQYQVNRGEPNN-----VGEEDCAEFGS 361
QY 288 DGRNDDVQCRPYHWVCEAGLGQTSQE 314
Db 362 NG-WNDDKCNLAKFWICKKSAASCSRD 387

RESULT 11

JC7595

Scavenger receptor with C-type lectin type I - human

C;Species: Homo sapiens (man)

C;Date: 30-Jun-2001 #sequence_revision 30-Jun-2001 #text_change 09-Jul-2004

C;Accession: JC7595

R;Nakamura, K.; Funakoshi, H.; Miyamoto, K.; Tokunaga, F.; Nakamura, T.

Biochem. Biophys. Res. Commun. 280, 1028-1035, 2001

A;Title: Molecular cloning and functional characterization of a human scavenger receptor

A;Reference number: JC7595; MUID:21092718; PMID:11162630

A;Contents: Placenta

A;Accession: JC7595

A;Molecule type: mRNA

A;Residues: 1-742 <NAK>

A;Cross-references: UNIPROT:Q9BYH7; DDBJ:AB038518

C;Comment: This receptor, a member of the scavenger receptor family, belonging to the tyrosine kinase receptor subfamily, plays a role in recognizing and internalizing

C;Genetics:

A;Gene: srcl-I

A;Map position: 18p11.32

C;Keywords: coiled coil; glycoprotein; transmembrane protein

F;1-39/Domain: cytosolic (amino-terminus) #status predicted <CYT>

F;16-19/Region: internalization signal YKRF

F;40-56/Domain: transmembrane #status predicted <TM>

F;57-112/Domain: extracellular #status predicted <EXT>

F;113-335/Domain: coiled coil #status predicted <COC>

F;369-384/Region: serine/threonine-rich #status predicted

F;443-589/Domain: collagen-like #status predicted <COL>

F;607-732/Domain: C-type lectin/carbohydrate recognition domain #status predicted <CRD>

A>Note: residues 24-48 form an uncharged, hydrophobic region that may interact with or e
 J.Mellow, T.E.; Halberg, D.; Drickamer, K.
 J. Biol. Chem. 263, 5468-5473, 1988
 A>Title: Endocytosis of N-acetylglucosamine-containing glycoproteins by rat fibroblasts
 A:Reference number: A28194; MUID:88186849; PMID:3281941
 A:Accession: A28194
 A:Molecule type: mRNA
 A:Residues: 1-207 <HEL>
 A:CROSS-references: GB:J03188; NID:9212246; PIDN:AAA48937.1; PID:g212247
 R:Bezouska, K.; Crichtlow, G.V.; Rose, J.M.; Taylor, M.E.; Drickamer, K.
 J. Biol. Chem. 266, 11604-11609, 1991
 A>Title: Evolutionary conservation of intron position in a subfamily of genes encoding c
 A:Reference number: A40427; MUID:91268022; PMID:2050668
 A:Accession: A40427
 A:Molecule type: DNA
 A:Residues: 1-207 <BEZ>
 A:CROSS-references: GB:M63225; GB:M63226; GB:M63227; GB:M63228; GB:M63229; GB:M63230; NI
 A:CROSS-references: GB:M63225; GB:M63226; GB:M63227; GB:M63228; GB:M63229; GB:M63230; NI
 C:Comment: Hepatic lectin is a membrane receptor protein that recognizes and binds expos
 and endocytosis.
 C:Genetics:
 A:Introns: 15/1; 50/1; 75/1; 125/3; 163/2
 A:Superfamily: Hepatic lectin; C-type lectin homology
 C:Keywords: acetylated amino end; glycoprotein; lectin; transmembrane protein
 F:1-23/Domain: Intracellular #status predicted <INT>
 F:24-47/Domain: transmembrane #status predicted <TEA>
 F:48-207/Domain: extracellular #status predicted <EXT>
 F:78-201/Domain: C-type lectin homology <LCH>
 F:1/Modified site: acetylated amino end (Met) #status experimental
 F:67/Binding site: carbohydrate (Asn) (covalent) #status experimental
 Query Match 16.2%; Score 281; DB 1; Length 207; \\
 Best Local Similarity 36.7%; Pred. No. 1.6e-15;
 Matches 62; Conservative 26; Mismatches 67; Indels 14; Gaps 4;
 Db 147 SEMLLRVQQLVQDLKKLTQVATLNNASPEGRCCPV-----NWVHQDSCYWFSSHGMS 201
 44 SVSLARIAALSSKLS--TLQSEPKHFNFSRDSLLFCGASQSQWYFEGRCYFSLRMS 101
 QY 202 WBAEAKYCOLKNAHLVINSREPNFQVYKLSAYTWMGLSD--PEGAKWVDGTDYATG 259
 Db 102 WHKAKAECEMHSLIIDSYAKQNFVFRTRNERFWIGLTDENQGEQWQVVDGTDTRSS 161
 QY 260 FQWKWQGDQDDWQGHGLGGEDCAHFPDGRWNDVQCRPHYVWCAGL 308
 Db 162 FTFWKEGEPNN-----RGFNEDCAHYWTSGQWMDVYCTCYVYVCEKPL 205
 RESULT 15
 LNHRU
 IGE Fc receptor II, low-affinity [validated] - human
 N:Alternate names: Blast-2; CD23; Fc-epsilon-R-II; lymphocyte IGE receptor
 N:Contains: IGE Fc receptor II, splice form a; IGE Fc receptor II, splice form a'; IGE F
 C:Species: Homo sapiens (man)
 C:Date: 31-Mar-1988 #sequence revision 31-Mar-1988 #text change 09-Jul-2004
 C:Accession: A26067; S03279; S39442; S39443; A26164; A26589; A31924; JLO132; S29107
 R:Kikutani, H.; Inui, S.; Sato, R.; Barsumian, E.L.; Owaki, H.; Yamasaki, K.; Kaisho, T.
 Cell 47, 657-665, 1986
 A>Title: Molecular structure of human lymphocyte receptor for immunoglobulin E.
 A:Reference number: A26067; MUID:87051737; PMID:2877743
 A:Accession: A26067
 A:Molecule type: mRNA
 A:Residues: 1-321 <KIK>
 A:CROSS-references: UNIPROT:P06734; GB:M14766; NID:g182449; PIDN:AAA52435.1; PID:g182450
 A:Experimental source: EBV-transformed B lymphoblastoid cells RPMI-8866
 R:Suter, U.; Bastos, R.; Hofstetter, H.
 Nucleic Acids Res. 15, 7295-7308, 1987
 A>Title: Molecular structure of the gene and the 5'-flanking region of the human lymphoc
 A:Reference number: S03279; MUID:88015536; PMID:2958779
 A:Accession: S03279
 A:Status: nucleic acid sequence not shown; not compared with conceptual translation
 A:Molecule type: DNA
 A:Residues: 157-284 <SUT>
 A:CROSS-references: GB:X06049; NID:g31316

A>Note: all exon sequences were determined but the complete sequence is not shown
 R.Matsui, M.; Nunez, R.; Sachi, Y.; Lynch, R.G.; Yodoi, J.
 FEBS Lett. 335, 51-56, 1993
 A>Title: Alternative transcripts of the human CD23/Fc-epsilon-R-II. A possible novel mech
 A:Reference number: S39442; MUID:94063078; PMID:8243664
 A:Accession: S39442
 A:Molecule type: DNA
 A:Residues: 1-7, 'D', 47-50 <MAS1>
 A:Experimental source: splice form a'
 A:Accession: S39443
 A:Molecule type: DNA
 A:Residues: 'MNPPSD', 47-50 <MAS2>
 A:Experimental source: splice form b'
 R:Judin, C.; Hofstetter, H.; Sarfati, M.; Levy, C.A.; Suter, U.; Alaimo, D.; Kilcherr, I.
 EMBO J. 6, 109-114, 1987
 A>Title: Cloning and expression of the cDNA coding for a human lymphocyte IGE receptor.
 A:Reference number: A26164; MUID:87218454; PMID:3034567
 A:Accession: A26164
 A:Molecule type: mRNA
 A:Residues: 1-288, 'T', 270-321 <LUD>
 A:CROSS-references: GB:X04772; NID:g34002; PIDN:CAA28465.1; PID:g34003
 A>Note: the codon given for 269-Asn (ACC) is inconsistent with the authors' translation
 A>Note: part of this sequence, including the amino end of soluble forms of the protein, v
 R:Kuka, K.; Takami, M.; Kim, C.W.; Honjo, T.; Miyoshi, T.; Tagaya, Y.; Kawabe, T.; Yodoi,
 Proc. Natl. Acad. Sci. U.S.A. 84, 819-823, 1987
 A>Title: Human lymphocyte Fc receptor for IGE: sequence homology of its cloned cDNA with
 A:Reference number: A26589; MUID:87118255; PMID:2949326
 A:Accession: A26589
 A:Molecule type: mRNA
 A:Residues: 1-321 <IKU>
 A:CROSS-references: GB:M15059; NID:g182447; PIDN:AAA52434.1; PID:g182448
 A>Note: part of this sequence, including the amino end of soluble forms of the protein, v
 R:Yokota, A.; Kikurani, H.; Tanaka, T.; Sato, R.; Barsumian, E.L.; Suenura, M.; Kishimoc
 Cell 55, 611-618, 1988
 A>Title: Two species of human Fc-epsilon receptor II (Fc-epsilon-R-II/CD23): tissue-specifi
 A:Reference number: A31924; MUID:89028672; PMID:2972386
 A:Accession: A31924
 A:Molecule type: mRNA
 A:Residues: 'MNPPSQ', 8-14 <YOK>
 A:CROSS-references: GB:M23562; NID:g182444
 A:Experimental source: splice form IIB
 R:Letellier, M.; Sarfati, M.; Delespesse, G.
 Mol. Immunol. 26, 1105-1112, 1989
 A>Title: Mechanisms of formation of IGE-binding factors (soluble CD23)-I. Fc epsilon R II
 A:Reference number: JLO132; MUID:90220658; PMID:2534424
 A:Accession: JLO132
 A:Molecule type: protein
 A:Residues: 1-321 <LUT>
 A:Experimental source: lymphoblastoid B cell line
 R:Rose, K.; Turcatti, G.; Graber, P.; Pochon, S.; Regamey, P.O.; Jansen, K.U.; Magnenat,
 Biochem. J. 286, 819-824, 1992
 A>Title: Partial characterization of natural and recombinant human soluble CD23.
 A:Reference number: S29107; MUID:93038513; PMID:1417742
 A:Accession: S29107
 A:Molecule type: protein
 A:Residues: 152-166;173-179;189-212;230-263;268-306 <ROS>
 R:Padan, E.A.; Helm, B.A.
 submitted to the Brookhaven Protein Data Bank, June 1993
 A:Reference number: A51791; PDB:1HLI
 A:Contents: annotation; conformation by theoretical model, residues 173-285
 R:Sajorath, J.
 submitted to the Brookhaven Protein Data Bank, November 1995
 A:Reference number: A65963; PDB:1KJE
 A:Contents: annotation; conformation by theoretical model, residues 173-285
 C:Comment: The sequence of the splice form a is shown.
 C:Comment: This receptor for the Fc portion of IGE is expressed in various hematopoietic
 cells.
 C:Comment: Splice form a is expressed constitutively in B-cells; b is expressed in other
 C:Comment: Soluble IGE-binding factors are produced by proteolytic cleavage of IGE Fc re
 C:Genetics:
 A:Gene: GDB:FCER2; FCE2
 A:CROSS-references: GDB:118888; OMIM:151445
 A:Map position: 19p13.3-19p13.3

A; Introns: 8/1; 46/1; 64/1; 85/1; 106/1; 127/1; 157/1; 207/3; 243/2	
C; Superfamily: IGE receptor II; C-type lectin homology	
C; Keywords: alternative splicing; B-cell; glycoprotein; immunoglobulin receptor; macroph	
F; 'MNPPSQ', 47-321/Product: IGE Fc receptor II, splice form b' #status predicted <SFBI>	
F; 'MNPPSQ', 8-321/Product: IGE Fc receptor II, splice form b #status predicted <SFB>	
F; 1-321/Product: IGE Fc receptor II, splice form a #status predicted <SFA>	
F; 1-23/Domain: intracellular #status predicted <INT>	
F; 1-7, 'D', 47-321/Product: IGE Fc receptor II, splice form a' #status predicted <SFAI>	
F; 14-20/Region: stop-transfer sequence	
F; 22-45/Domain: transmembrane #status predicted <TM>	
F; 46-321/Domain: extracellular #status predicted <EXT>	
F; 64-84/Region: 21-residue repeat	
F; 81-321/Product: soluble IGE-binding factor (37K) #status predicted <IGE>	
F; 85-105/Region: 21-residue repeat	
F; 102-321/Product: soluble IGE-binding factor (33K) #status predicted <IGB>	
F; 106-126/Region: 21-residue repeat	
F; 125-321/Product: soluble IGE-binding factor (29K) #status predicted <IGI>	
F; 148-321/Product: soluble IGE-binding factor (25-27K), long form #status experimental <	
F; 150-321/Product: soluble IGE-binding factor (25-27K), short form #status experimental <	
F; 163-282/Domain: C-type lectin homology <LCH>	
F; 63/Binding site: carbohydrate (Asn) (covalent) #status predicted	
F; 147-148/Cleavage site: Lys-Leu (unidentified proteinase) #status experimental	
F; 149-150/Cleavage site: Arg-Met (unidentified proteinase) #status experimental	
F; 191-282, 259-273/Disulfide bonds: #status experimental	
Query Match 15.8%; Score 274.5; DB 1; Length 321;	
Best Local Similarity 29.4%; Pred. No. 9.4e-15;	
Matches 74; Conservative 32; Mismatches 87; Indels 59; Gaps 10;	
Qy	61 QNSKFORDLVTLTDFSNFTSNVAIEIQALTSQGSLEETIASLKAIEVGFQERQAGVS 120
Db	84 QSTQISQLEELR-----AQQRLLKSQDLSELSWNLGLQADLSSFKSQ----- 126
Qy	121 ELQEHHTQKAHLGHCPFCVCPVHSEMLLRVQQLVQDLKKLTQCVATLNNASTEGTC 180
Db	127 ELNERNEA-----SDLLERLRREVTKLKLR-----MELQVSSGFVC 160
Qy	181 --CPVNWVEHQDSQCYWFHSHGMSWAEAEKYCOLKNAHLVIVINSREBNFVQKYLGSAYTW 238
Db	161 NTCPEKWINFQRCYFYFGKGTQKWVHARYACDDMEGQLVSIHSPEDFLTKHASHTGSW 220
Qy	239 MGLS--DPGAMKWVDGT--DYATGFQNWKPQPDMDQGHGLGGGDECAHFHPDGRWNDD 294
Db	221 IGLRNLDLKGEFTWVDGSHVDYS-----NNAPEPT-----SRSQGEDCYVMRSGRWNDA 271
Qy	295 VCORPY-HWVCE 305
Db	272 FCORLGAWVCD 283

Search completed: May 27, 2005, 12:49:42
Job time : 19 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 27, 2005, 12:44:49 ; Search time 143 Seconds
(without alignments)
762.317 Million cell updates/sec

Title: US-10-829-107-4

Perfect score: 1738

Sequence: 1 MTRTYENFQYLENKVKVQGF.....QRPYHWCAGLGQTSQESH 316

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

*Searched: 1462099 seqs, 344972447 residues

Total number of hits satisfying chosen parameters: 1462099

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/prodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/prodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/prodata/2/pubpaa/US06_PUBCOMB.pep.*
4: /cgn2_6/prodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/prodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/prodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/prodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/prodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/prodata/2/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/prodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/prodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/prodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/prodata/2/pubpaa/US10_PUBCOMB.pep.*
14: /cgn2_6/prodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/prodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/prodata/2/pubpaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/prodata/2/pubpaa/US10_NEW_PUB.pep.*
18: /cgn2_6/prodata/2/pubpaa/US11_NEW_PUB.pep.*
19: /cgn2_6/prodata/2/pubpaa/US60_NEW_PUB.pep.*
20: /cgn2_6/prodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1738	100.0	316	9	US-09-862-802-4
2	1738	100.0	316	16	US-10-829-107-4
3	1718.5	98.9	319	15	US-10-138-588-90
4	1550.5	89.2	289	16	US-10-829-107-13
5	1544	88.8	292	16	US-10-829-107-12
6	1427.5	82.1	273	9	US-09-862-802-10
7	1427.5	82.1	273	16	US-10-829-107-10
8	1410.5	81.2	267	15	US-10-138-588-88
9	923.5	53.1	291	9	US-09-862-802-5
10	923.5	53.1	291	16	US-10-829-107-5
11	783	45.1	287	9	US-09-862-802-6
12	783	45.1	287	15	US-10-236-392-10
13	783	45.1	287	16	US-10-829-107-6

14	760	43.7	306	15	US-10-236-392-8	Sequence 8, Appli
15	754.5	43.4	311	15	US-10-236-392-6	Sequence 6, Appli
16	326.5	18.8	550	15	US-10-072-012-513	Sequence 513, App
17	321	18.5	345	15	US-10-262-839-94	Sequence 94, Appl
18	316	18.2	404	10	US-09-891-894-1	Sequence 1, Appli
19	316	18.2	404	14	US-10-151-274-2	Sequence 2, Appli
20	316	18.2	404	14	US-10-184-150-1	Sequence 1, Appli
21	316	18.2	404	15	US-10-369-214-131	Sequence 131, App
22	316	18.2	404	15	US-10-328-997-1	Sequence 1, Appli
23	313	18.0	342	14	US-10-223-085-28	Sequence 28, Appl
24	313	18.0	342	14	US-10-223-084-28	Sequence 28, Appl
25	313	18.0	342	14	US-10-223-088-28	Sequence 28, Appl
26	313	18.0	342	14	US-10-223-090-28	Sequence 28, Appl
27	313	18.0	342	14	US-10-223-087-28	Sequence 28, Appl
28	313	18.0	342	14	US-10-223-083-28	Sequence 28, Appl
29	313	18.0	342	14	US-10-223-089-28	Sequence 28, Appl
30	313	18.0	342	14	US-10-223-081-28	Sequence 28, Appl
31	313	18.0	342	14	US-10-223-082-28	Sequence 28, Appl
32	313	18.0	342	15	US-10-305-654-28	Sequence 28, Appl
33	313	18.0	342	15	US-10-081-056-28	Sequence 28, Appl
34	313	18.0	457	9	US-09-745-763-218	Sequence 218, App
35	313	18.0	542	9	US-09-745-763-219	Sequence 219, App
36	313	18.0	618	14	US-10-203-860-24	Sequence 24, Appl
37	313	18.0	742	14	US-10-308-448-11	Sequence 11, Appl
38	311	17.9	399	15	US-10-451-459-2	Sequence 2, Appli
39	311	17.9	742	14	US-10-203-860-2	Sequence 2, Appli
40	311	17.9	742	15	US-10-341-434-85	Sequence 85, Appl
41	309.5	17.8	546	15	US-10-072-012-158	Sequence 158, App
42	309.5	17.8	548	15	US-10-072-012-512	Sequence 512, App
43	309.5	17.8	549	15	US-10-072-012-160	Sequence 160, App
44	301	17.3	399	10	US-09-891-894-2	Sequence 2, Appli
45	301	17.3	399	14	US-10-184-150-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-09-862-802-4
; Sequence 4, Application US/09862802
; Patent No. US20020165346A1
; GENERAL INFORMATION:
; APPLICANT: Schering-Plough Corporation
; TITLE OF INVENTION: MAMMALIAN MEMBRANE PROTEIN GENES; RELATED REAGENTS
; FILE REFERENCES: SF06958
; CURRENT APPLICATION NUMBER: US/09/862,802
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 09/111,470
; PRIOR FILING DATE: 1998-07-08
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 4
; LENGTH: 316
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: mammalian protein
US-09-862-802-4

Query Match	100.0%;	Score	1738;	DB	9;	Length	316;
Best Local Similarity	100.0%;	Pred. No.	3.1e-151;				
Matches	316;	Conservative	0;	Mismatches	0;	Indels	0;
Gaps	0;						
Qy	1	MTRTYENFQYLENKVKVQGF	PLQSLQLRLSGPCHLLLSGLGLLLVTCVVG	60			
Db	1	MTRTYENFQYLENKVKVQGF	PLQSLQLRLSGPCHLLLSGLGLLLVTCVVG	60			
Qy	61	QNSKQFDLVTLRDTSNFTSV	IAEIQALTSQGSLEETIASLKAEVEGFKQERQ	120			
Db	61	QNSKQFDLVTLRDTSNFTSV	IAEIQALTSQGSLEETIASLKAEVEGFKQERQ	120			
Qy	121	ELOEHTTKAHLCCHPCSV	CVVHSEMLLRVQOLVODLKLTCQVATLNNAST	180			
Db	121	ELOEHTTKAHLCCHPCSV	CVVHSEMLLRVQOLVODLKLTCQVATLNNAST	180			

Db 121 ELQEHHTQKAHLGHCHPCPSVCVPVHSEMLLRVQQLVQDLKKLTCCQVATLNNASTEGTC 180
QY 181 CPVNWVEHODSCYWFSGMSWAEAEKYCOLKNAHLVWVINSREONFVKYLGSAYTWMG 240
Db 181 CPVNWVEHODSCYWFSGMSWAEAEKYCOLKNAHLVWVINSREONFVKYLGSAYTWMG 240
QY 241 LSDPEGAKWVDGTDYATGFQNWKPQDDWQGHGLGGGEDCAHFHPDGRWDDVCORPY 300
Db 241 LSDPEGAKWVDGTDYATGFQNWKPQDDWQGHGLGGGEDCAHFHPDGRWDDVCORPY 300
QY 301 HWVCEAGLGQTSQESH 316
Db 301 HWVCEAGLGQTSQESH 316

RESULT 2

US-10-829-107-4
; Sequence 4, Application US/10829107
; Publication No. US20040192892A1
; GENERAL INFORMATION:
; APPLICANT: Valladeau, Jenny
; APPLICANT: Ravel, Odile
; APPLICANT: Bates, Elizabeth Ester Mary
; APPLICANT: Ford, John
; APPLICANT: Lebecque, Serge J.E.
; APPLICANT: Saeland, Sem
; TITLE OF INVENTION: Isolated Mammalian Membrane Protein Genes; Related Reagents
; FILE REFERENCE: SF0695 B
; CURRENT APPLICATION NUMBER: US/10/829,107
; CURRENT FILING DATE: 2004-04-21
; PRIOR APPLICATION NUMBER: US/09/862,802A
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/053,080
; PRIOR FILING DATE: 1997-07-09
; PRIOR APPLICATION NUMBER: US 09/111,470
; PRIOR FILING DATE: 1998-07-08
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 316
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: mammalian nucleic acid and protein
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (608)..(673)
; OTHER INFORMATION: short form lacks nucleotides 608-673
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (775)..(776)
; OTHER INFORMATION: ASGPRM (Table 2) has sequence insert encoding GEE between nucleot
; OTHER INFORMATION: ides 775-776
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1064)..(1064)
; OTHER INFORMATION: nucleotide 1064 of DCM28 may be A, which would encode Asn rather
; OTHER INFORMATION: than Asp at the residue numbered 270
US-10-829-107-4

Query Match 100.0%; Score 1738; DB 16; Length 316;
Best Local Similarity 100.0%; Pred. No. 3.1e-151;
Matches 316; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTRTYENFQYLENKVKVQGFKNGLPLQSLQRLRSGPCHLLLSLGLGLLLVLCVVG 60
Db 1 MTRTYENFQYLENKVKVQGFKNGLPLQSLQRLRSGPCHLLLSLGLGLLLVLCVVG 60
QY 61 QNSKFORDLVLTDFSNFTSNVTVAEIQALTSQSSLEETIASLKAEEVGFQKQAGVS 120
Db 61 QNSKFORDLVLTDFSNFTSNVTVAEIQALTSQSSLEETIASLKAEEVGFQKQAGVS 120
QY 121 ELQEHHTQKAHLGHCHPCPSVCVPVHSEMLLRVQQLVQDLKKLTCCQVATLNNASTEGTC 180

Db 121 ELQEHHTQKAHLGHCHPCPSVCVPVHSEMLLRVQQLVQDLKKLTCCQVATLNNASTEGTC 180
QY 181 CPVNWVEHODSCYWFSGMSWAEAEKYCOLKNAHLVWVINSREONFVKYLGSAYTWMG 240
Db 181 CPVNWVEHODSCYWFSGMSWAEAEKYCOLKNAHLVWVINSREONFVKYLGSAYTWMG 240
QY 241 LSDPEGAKWVDGTDYATGFQNWKPQDDWQGHGLGGGEDCAHFHPDGRWDDVCORPY 300
Db 241 LSDPEGAKWVDGTDYATGFQNWKPQDDWQGHGLGGGEDCAHFHPDGRWDDVCORPY 300
QY 301 HWVCEAGLGQTSQESH 316
Db 301 HWVCEAGLGQTSQESH 316

RESULT 3

US-10-138-588-90
; Sequence 90, Application US/10138588
; Publication No. US20040018594A1
; GENERAL INFORMATION:
; APPLICANT: Alsobrook et al.
; TITLE OF INVENTION: NOVEL ANTIBODIES THAT BIND TO ANTIGENIC POLYPEPTIDES, NUCLEIC ACI
; FILE REFERENCE: 21402-347A
; CURRENT APPLICATION NUMBER: US/10/138,588
; CURRENT FILING DATE: 2002-05-01
; PRIOR APPLICATION NUMBER: 60/288,395
; PRIOR FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: 60/308,901
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/313,388
; PRIOR FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: 60/324,757
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: 60/288,900
; PRIOR FILING DATE: 2001-05-04
; NUMBER OF SEQ ID NOS: 203
; SEQ ID NO 90
; LENGTH: 319
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-138-588-90

Query Match 98.9%; Score 1718.5; DB 15; Length 319;
Best Local Similarity 98.7%; Pred. No. 1.9e-149;
Matches 315; Conservative 0; Mismatches 1; Indels 3; Gaps 1;

QY 1 MTRTYENFQYLENKVKVQGFKNGLPLQSLQRLRSGPCHLLLSLGLGLLLVLCVVG 60
Db 1 MTRTYENFQYLENKVKVQGFKNGLPLQSLQRLRSGPCHLLLSLGLGLLLVLCVVG 60
QY 61 QNSKFORDLVLTDFSNFTSNVTVAEIQALTSQSSLEETIASLKAEEVGFQKQAGVS 120
Db 61 QNSKFORDLVLTDFSNFTSNVTVAEIQALTSQSSLEETIASLKAEEVGFQKQAGVS 120
QY 121 ELQEHHTQKAHLGHCHPCPSVCVPVHSEMLLRVQQLVQDLKKLTCCQVATLNN---ASTE 177
Db 121 ELQEHHTQKAHLGHCHPCPSVCVPVHSEMLLRVQQLVQDLKKLTCCQVATLNNNGEASTE 180
QY 178 GTCCPVNWVEHODSCYWFSGMSWAEAEKYCOLKNAHLVWVINSREONFVKYLGSAYT 237
Db 181 GTCCPVNWVEHODSCYWFSGMSWAEAEKYCOLKNAHLVWVINSREONFVKYLGSAYT 240
QY 238 WMGLSDPEGAKWVDGTDYATGFQNWKPQDDWQGHGLGGGEDCAHFHPDGRWDDVCQ 297
Db 241 WMGLSDPEGAKWVDGTDYATGFQNWKPQDDWQGHGLGGGEDCAHFHPDGRWDDVCQ 300
QY 298 RPYHWVCEAGLGQTSQESH 316
Db 301 RPYHWVCEAGLGQTSQESH 319


```
QY 1 MTRTYENFOYLENKVKVQGFQKNGPLQSLQRLRSGPCHLLSLGLGLLLVLIICVVG 60
Db 1 MTRTYENFOYLENKVKVQGFQKNGPLQSLQRLRSGPCHLLSLGLGLLLVLIICVVG 41
QY 61 QNSKFORDLVTLRTDTSNFTSNVTVAIQALTSQSSLEETIASLKAEVEGFKQEROAGVS 120
Db 42 QNSKFORDLVTLRTDTSNFTSNVTVAIQALTSQSSLEETIASLKAEVEGFKQEROA--- 98
QY 121 ELQEHHTQKAHLGCHPCFVCVVPVHSEMILRRVQQLVQDLKKLTCCQVATLNN---ASTE 177
Db 99 -----VHSEMILRRVQQLVQDLKKLTCCQVATLNNGBEASTE 134
QY 178 GTCCPVNVWEHODSCYWFHSGMSWAEAEKYCOLKNAHLVWVINSREEQNFVKYILGSAYT 237
Db 135 GTCCPVNVWEHODSCYWFHSGMSWAEAEKYCOLKNAHLVWVINSREEQNFVKYILGSAYT 194
QY 238 WMGLSDPEGAWKWVDGTDYATGFQNWKPQDDWQHGGLGGEDCAHFHPDGRWDDVQC 297
Db 195 WMGLSDPEGAWKWVDGTDYATGFQNWKPQDDWQHGGLGGEDCAHFHPDGRWDDVQC 254
QY 298 RPYHWVCEAGLGQTSQESH 316
Db 255 RPYHWVCEAGLGQTSQESH 273

RESULT 7
US-10-829-107-10
; Sequence 10, Application US/10829107
; Publication No. US20040192892A1
; GENERAL INFORMATION:
; APPLICANT: Valladeau, Jenny
; APPLICANT: Ravel, Odile
; APPLICANT: Bates, Elizabeth Ester Mary
; APPLICANT: Ford, John
; APPLICANT: Lebecque, Serge J.E.
; APPLICANT: Saeland, Sem
; TITLE OF INVENTION: Isolated Mammalian Membrane Protein Genes; Related Reagents
; FILE REFERENCE: SF0695 B
; CURRENT APPLICATION NUMBER: US/10/829,107
; PRIOR FILING DATE: 2004-04-21
; PRIOR APPLICATION NUMBER: US/09/862,802A
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/053,080
; PRIOR FILING DATE: 1997-07-09
; PRIOR APPLICATION NUMBER: US 09/111,470
; PRIOR FILING DATE: 1998-07-08
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 10
; LENGTH: 273
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: mammalian nucleic acid and protein
US-10-829-107-10

Query Match 82.1%; Score 1427.5; DB 16; Length 273;
Best Local Similarity 84.6%; Pred. No. 9.3e-123;
Matches 270; Conservative 0; Mismatches 0; Indels 49; Gaps 3;

QY 1 MTRTYENFOYLENKVKVQGFQKNGPLQSLQRLRSGPCHLLSLGLGLLLVLIICVVG 60
Db 1 MTRTYENFOYLENKVKVQGFQKNGPLQSLQRLRSGPCHLLSLGLGLLLVLIICVVG 41
QY 61 QNSKFORDLVTLRTDTSNFTSNVTVAIQALTSQSSLEETIASLKAEVEGFKQEROAGVS 120
Db 42 QNSKFORDLVTLRTDTSNFTSNVTVAIQALTSQSSLEETIASLKAEVEGFKQEROA--- 98
QY 121 ELQEHHTQKAHLGCHPCFVCVVPVHSEMILRRVQQLVQDLKKLTCCQVATLNN---ASTE 177
Db 99 -----VHSEMILRRVQQLVQDLKKLTCCQVATLNNGBEASTE 134
QY 178 GTCCPVNVWEHODSCYWFHSGMSWAEAEKYCOLKNAHLVWVINSREEQNFVKYILGSAYT 237
```

```
Db 135 GTCCPVNVWEHODSCYWFHSGMSWAEAEKYCOLKNAHLVWVINSREEQNFVKYILGSAYT 194
QY 238 WMGLSDPEGAWKWVDGTDYATGFQNWKPQDDWQHGGLGGEDCAHFHPDGRWDDVQC 297
Db 195 WMGLSDPEGAWKWVDGTDYATGFQNWKPQDDWQHGGLGGEDCAHFHPDGRWDDVQC 254
QY 298 RPYHWVCEAGLGQTSQESH 316
Db 255 RPYHWVCEAGLGQTSQESH 273

RESULT 8
US-10-138-588-88
; Sequence 88, Application US/10138588
; Publication No. US20040018594A1
; GENERAL INFORMATION:
; APPLICANT: Alsobrook et al.
; TITLE OF INVENTION: NOVEL ANTIBODIES THAT BIND TO ANTIGENIC POLYPEPTIDES, NUCLEIC ACT
; FILE REFERENCE: 21402-347A
; CURRENT APPLICATION NUMBER: US/10/138,588
; PRIOR FILING DATE: 2002-05-01
; PRIOR APPLICATION NUMBER: 60/288,395
; PRIOR FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: 60/308,901
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/313,388
; PRIOR FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: 60/324,757
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: 60/288,900
; PRIOR FILING DATE: 2001-05-04
; NUMBER OF SEQ ID NOS: 203
; SEQ ID NO 88
; LENGTH: 267
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-138-588-88

Query Match 81.2%; Score 1410.5; DB 15; Length 267;
Best Local Similarity 89.5%; Pred. No. 3.3e-121;
Matches 263; Conservative 0; Mismatches 4; Indels 27; Gaps 1;

QY 23 GPLPLQSLQRLRSGPCHLLSLGLGLLLVLIICVGFQNSKFORDLVTLRTDTSNFTSN 82
Db 1 GPLPLQSLQRLRSGPCHLLSLGLGLLLVLIICVGFQNSKFORDLVTLRTDTSNFTSN 60
QY 83 TVAEIQALTSQSSLEETIASLKAEVEGFKQEROAGVSELOEHTTQKAHLGCHPCPSVC 142
Db 61 TVAEIQALTSQSSLEETIASLKAEVEGFKQEROA----- 95
QY 143 VPVHSEMILRRVQQLVQDLKKLTCCQVATLNNASTEGTCCPVNVWEHODSCYWFHSGMSW 202
Db 96 --VHSEMILRRVQQLVQDLKKLTCCQVATLNNASTEGTCCPVNVWEHODSCYWFHSGMSW 153
QY 203 AEAEKYCOLKNAHLVWVINSREEQNFVKYILGSAYTWMGLSDPEGAWKWVDGTDYATGFQ 262
Db 154 AEAEKYCOLKNAHLVWVINSREEQNFVKYILGSAYTWMGLSDPEGAWKWVDGTDYATGFQ 213
QY 263 WKPGQDDWQHGGLGGEDCAHFHPDGRWDDVQCQRPYHWVCEAGLGQTSQESH 316
Db 214 WKPGQDDWQHGGLGGEDCAHFHPDGRWDDVQCQRPYHWVCEAGLGQTSQESH 267

RESULT 9
US-09-862-802-5
; Sequence 5, Application US/09862802
; Patent No. US20020165346A1
; GENERAL INFORMATION:
; APPLICANT: Schering-Plough Corporation
; TITLE OF INVENTION: MAMMALIAN MEMBRANE PROTEIN GENES; RELATED REAGENTS
; FILE REFERENCE: SF0695B
```

```

; CURRENT APPLICATION NUMBER: US/09/862,802
; CURRENT FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 09/111,470
; PRIOR FILING DATE: 1998-07-08
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 291
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: mammalian protein
US-09-862-802-5

Query Match
Best Local Similarity 53.1%; Score 923.5; DB 9; Length 291;
Matches 172; Conservative 47; Mismatches 67; Indels 29; Gaps 5;

QY 1 MTRTYENFOYLENKVK-VQGFKNGLPLOSLLQRLSRGCPCHLLSLGLGLLLVLCVWG 59
Db 1 MTKEYQDLQHLNDESDHQLRGKPPPPQPLLQRLCSGPRLLLSLGLSLLLVVVCVIG 60
QY 60 FQNSKFORDLVLTDFSNFTSNVTVAEIQALTSQGSLEETIASLKAEEVGFQKQROAGV 119
Db 61 SQNSQLQEBELRGURETFSNFTASTAQVKGSLTQGGNVGRKMKLSQLE--KQK---- 114
QY 120 SELOEHTTQKAHLGCHPCPSVCPVHSEMMLRVQQLVQDLKKLTQVATLNNNASTEGT 179
Db 115 -DLSED-----HSSLLLVHKQFVSDRLSRSLSCMAALQNGS--ERT 152
QY 180 CCPNVNVEHQDSCYWFSHSGMSWAEAEKYCOLKNAHLVWINSREEQNFVQKYLGSAYTWM 239
Db 153 CCPNVNVEHRSYWFSSRSKAWADADNYCRLEDAHLVWVTSWEEKFVQHHIGPVNTWM 212
QY 240 GLSDPEGAKWVDGTDYATGFQNWKPQDDWQGHGLGGEDCAHFHPDGRWDDVCORP 299
Db 213 GLHDQNGPKWVDGTDYETGFKWVRPEQPDWYGHGLGGEDCAHFTDDGRWDDVCORP 272
QY 300 YHWVCEAGLGQTSQE 314
Db 273 YRWVCETELDKASQE 287

RESULT 11
US-09-862-802-6
; Sequence 6, Application US/09862802
; Patent No. US20020165346A1
; GENERAL INFORMATION:
; APPLICANT: Schering-Plough Corporation
; TITLE OF INVENTION: MAMMALIAN MEMBRANE PROTEIN GENES; RELATED REAGENTS
; FILE REFERENCE: SF0695B
; CURRENT APPLICATION NUMBER: US/09/862,802
; CURRENT FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 09/111,470
; PRIOR FILING DATE: 1998-07-08
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 287
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: mammalian protein
US-09-862-802-6

Query Match
Best Local Similarity 45.1%; Score 783; DB 9; Length 287;
Matches 146; Conservative 50; Mismatches 89; Indels 28; Gaps 3;

QY 1 MTRTYENFOYLENKVKVQGFKNGLPLOSLLQRLSRGCPCHLLSLGLGLLLVLCVWG 60
Db 1 MAXDFQDIQQLSSEENDHPHQGPPPAQPLAQLRCSMVCPSLLALSFNILLVVICVGS 60
QY 61 QNSKFORDLVLTDFSNFTSNVTVAEIQALTSQGSLEETIASLKAEEVGFQKQROAGV 120
Db 61 QSAQLQALRSLKEAFSNFSSSTLTVEQAISTHGGSVGDKITSGLAKLEKQQQDLKA--- 117
QY 121 ELOEHTTQKAHLGCHPCPSVCPVHSEMMLRVQQLVQDLKKLTQVATLNNNASTEGT 180
Db 118 ---DHDLLEFLKHPF-----VDLRFVACQWELHSHNGS--ORTC 152
QY 181 CPNVNVEHQDSYWFSSHSGMSWAEAEKYCOLKNAHLVWINSREEQNFVQKYLGSAYTWM 240
Db 153 CPNVNVEHQDSYWFSSHSGMSWAEAEKYCOLKNAHLVWINSREEQNFVQHTNPTNWIG 212
QY 241 LSDPEGAKWVDGTDYATGFQNWKPQDDWQGHGLGGEDCAHFHPDGRWDDVCORP 300

; CURRENT APPLICATION NUMBER: US/09/862,802
; CURRENT FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 09/111,470
; PRIOR FILING DATE: 1998-07-08
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 291
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: mammalian protein
US-10-829-107-5

Query Match
Best Local Similarity 54.6%; Pred. No. 2.4e-76;
Matches 172; Conservative 47; Mismatches 67; Indels 29; Gaps 5;

QY 1 MTRTYENFOYLENKVK-VQGFKNGLPLOSLLQRLSRGCPCHLLSLGLGLLLVLCVWG 59
Db 1 MTKEYQDLQHLNDESDHQLRGKPPPPQPLLQRLCSGPRLLLSLGLSLLLVVVCVIG 60
QY 60 FQNSKFORDLVLTDFSNFTSNVTVAEIQALTSQGSLEETIASLKAEEVGFQKQROAGV 119
Db 61 SQNSQLQEBELRGURETFSNFTASTAQVKGSLTQGGNVGRKMKLSQLE--KQK---- 114
QY 120 SELOEHTTQKAHLGCHPCPSVCPVHSEMMLRVQQLVQDLKKLTQVATLNNNASTEGT 179
Db 115 -DLSED-----HSSLLLVHKQFVSDRLSRSLSCMAALQNGS--ERT 152
QY 180 CCPNVNVEHQDSCYWFSHSGMSWAEAEKYCOLKNAHLVWINSREEQNFVQKYLGSAYTWM 239
Db 153 CCPNVNVEHRSYWFSSRSKAWADADNYCRLEDAHLVWVTSWEEKFVQHHIGPVNTWM 212
QY 240 GLSDPEGAKWVDGTDYATGFQNWKPQDDWQGHGLGGEDCAHFHPDGRWDDVCORP 299
Db 213 GLHDQNGPKWVDGTDYETGFKWVRPEQPDWYGHGLGGEDCAHFTDDGRWDDVCORP 272
QY 300 YHWVCEAGLGQTSQE 314
Db 273 YRWVCETELDKASQE 287

RESULT 10
US-10-829-107-5
; Sequence 5, Application US/10829107
; Publication No. US20040192892A1
; GENERAL INFORMATION:
; APPLICANT: Valladeau, Jenny
; APPLICANT: Ravel, Odile
; APPLICANT: Bates, Elizabeth Ester Mary
; APPLICANT: Ford, John
; APPLICANT: Lebecque, Serge J.E.
; APPLICANT: Seland, Sem
; TITLE OF INVENTION: Isolated Mammalian Membrane Protein Genes; Related Reagents
; FILE REFERENCE: SF0695 B
; CURRENT APPLICATION NUMBER: US/10/829,107
; CURRENT FILING DATE: 2004-04-21
; PRIOR APPLICATION NUMBER: US/09/862,802A
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/053,080
; PRIOR FILING DATE: 1997-07-09
; PRIOR APPLICATION NUMBER: US 09/111,470
; PRIOR FILING DATE: 1998-07-08
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 291
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: mammalian protein
US-10-829-107-5
```



```
Db 61 QSAQLQAEALSKEAFNSFSSSTLTEVQAISTHGSGVDKITSIGAKLEKQOQDLKA--- 117
QY 121 ELQHTTQKAHLGCHPCSPVPHSEMLLRVOQLVQDLKULTCQVATLNNASTEGTC 180
Db 118 ---DHDALLFLHKFP-----VDLRFVACQWELLHNSGS-QRTC 152
QY 181 CPNVVVEHODSCYWFHSGMSWAEAEKYCOLKNAHLVWVINSREOFNVQKYLGSAYTMG 240
Db 153 CPNVVVEHODSCYWFHSGMSWAEAEKYCOLKNAHLVWVINSREOFNVQKYLGSAYTMG 212
QY 241 LSDPEGAKWVDGTDYATGQFNKPKQPDWQGHGGLGGEDCAHFHPDGRWDDVQCPY 300
Db 213 LTSDSGSKWVGDGTDYRHNKNAVTPQDNWHGHELGSSEDCVEVQPDGRWDDVQCPY 272
QY 301 HWVCEAGLGQTSQ 313
Db 273 RWVCEKRRNATGE 285

RESULT 14
US-10-236-392-8
; Sequence 8, Application US/10236392
; Publication No. US20040067490A1
; GENERAL INFORMATION:
; APPLICANT: Anderson, David W
; APPLICANT: Boldog, Ferenc L
; APPLICANT: Burgess, Catherine, E
; APPLICANT: Casman, Stacie J
; APPLICANT: Catterton, Elina
; APPLICANT: Chapoval, Andrei
; APPLICANT: Crabtree, Julie
; APPLICANT: Edinger, Shlomit, R
; APPLICANT: Ellerman, Karen
; APPLICANT: Gorman, Linda
; APPLICANT: Grosse, William M
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Li, Li
; APPLICANT: MacDougall, John R
; APPLICANT: Malyankar, Uriel M
; APPLICANT: Miller, Charles E
; APPLICANT: Millet, Isabelle
; APPLICANT: Padigar, Muralidhara
; APPLICANT: Patturajan, Meera
; APPLICANT: Pena, Carol A
; APPLICANT: Peyman, John A
; APPLICANT: Reiger, Daniel K
; APPLICANT: Rothenberg, Mark E
; APPLICANT: Shenoy, Suresh
; APPLICANT: Shimkets, Richard A
; APPLICANT: Smithson, Glenda
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-442A
; CURRENT FILING DATE: 2002-09-06
; PRIOR FILING DATE: 2002-03-30
; PRIOR APPLICATION NUMBER: US09/540,763
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: US60/390,155
; PRIOR FILING DATE: 2002-06-19
; PRIOR APPLICATION NUMBER: US09/635,949
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: US60/318,765
; PRIOR FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: US60/357,303
; PRIOR FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: US60/367,753
; PRIOR FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER: US60/369,479
; PRIOR FILING DATE: 2002-04-02
; PRIOR APPLICATION NUMBER: US09/659,634
```

```
; PRIOR FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: US60/318,120
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: US60/318,130
; PRIOR FILING DATE: 2001-09-07
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 794
; SOFTWARE: Custom
; SEQ ID NO 8
; LENGTH: 306
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-236-392-8

Query Match 43.7%; Score 760; DB 15; Length 306;
Best Local Similarity 48.3%; Pred. No. 2.8e-61;
Matches 142; Conservative 44; Mismatches 80; Indels 28; Gaps 3;

QY 20 FKNGPLQSLQLRSGPCHLLSLGLGLLVLVGVGFONSKFQRLVLTLPDSNF 79
Db 39 FLKGPPLQAPLAQRLCSMVCFSLLLSFILLVLLVICVTGSSQSAQLQAEALSKEAFSNF 98
QY 80 TSNTVAEIQALTSGSSLEETIASLKAEEVGFQKQAGVSEIQEHTTQKAHLGCHPCP 139
Db 99 SSTLTVEVQAISTHGSGVDKITSIGAKLEKQOQDLKA-----DHDALLFLHKFP--- 149
QY 140 SVCVPVHSEMLLRVOQLVQDLKULTCQVATLNNASTEGTCCPNVWVEHODSCYWFHSG 199
Db 150 -----VDLRFVACQWELLHNSGS-QRTCCPNVWVEHODSCYWFHSG 190
QY 200 MSWAEAEKYCOLKNAHLVWVINSREOFNVQKYLGSAYTMGLSDPEGAKWVDGTDYATG 259
Db 191 KAWAEAEKYCLLENHLVWVINSWEEQKFIVQHTNPFTWGLTSDSGSKWVGDGTDYRHN 250
QY 260 FQNWKPGQPDWQGHGGLGGEDCAHFHPDGRWDDVQCPYHWVCEAGLGQTSQ 313
Db 251 YKNWAVTPQDNWHGHELGSSEDCVEVQPDGRWDDVQCPYRWVCEKRRNATGE 304
```

```
RESULT 15
US-10-236-392-6
; Sequence 6, Application US/10236392
; Publication No. US20040067490A1
; GENERAL INFORMATION:
; APPLICANT: Anderson, David W
; APPLICANT: Boldog, Ferenc L
; APPLICANT: Burgess, Catherine, E
; APPLICANT: Casman, Stacie J
; APPLICANT: Catterton, Elina
; APPLICANT: Chapoval, Andrei
; APPLICANT: Crabtree, Julie
; APPLICANT: Edinger, Shlomit, R
; APPLICANT: Ellerman, Karen
; APPLICANT: Gerlach, Valerie
; APPLICANT: Gorman, Linda
; APPLICANT: Grosse, William M
; APPLICANT: Gusev, Vladamir
; APPLICANT: Kekuda, Ramesh
; APPLICANT: LaRoche, William J
; APPLICANT: Li, Li
; APPLICANT: MacDougall, John R
; APPLICANT: Malyankar, Uriel M
; APPLICANT: Miller, Charles E
; APPLICANT: Millet, Isabelle
; APPLICANT: Padigar, Muralidhara
; APPLICANT: Patturajan, Meera
; APPLICANT: Pena, Carol A
; APPLICANT: Peyman, John A
; APPLICANT: Rastelli, Luca
; APPLICANT: Reiger, Daniel K
; APPLICANT: Rothenberg, Mark E
; APPLICANT: Shenoy, Suresh
; APPLICANT: Shimkets, Richard A
```

```
/ APPLICANT: Smithson, Glenda
/ TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME
/ FILE REFERENCE: 21402-442A
/ CURRENT APPLICATION NUMBER: US/10/236,392
/ CURRENT FILING DATE: 2002-09-06
/ PRIOR APPLICATION NUMBER: US09/540,763
/ PRIOR FILING DATE: 2000-03-30
/ PRIOR APPLICATION NUMBER: US60/390,155
/ PRIOR FILING DATE: 2002-06-19
/ PRIOR APPLICATION NUMBER: US09/635,949
/ PRIOR FILING DATE: 2000-08-10
/ PRIOR APPLICATION NUMBER: US60/318,765
/ PRIOR FILING DATE: 2001-09-12
/ PRIOR APPLICATION NUMBER: US60/357,303
/ PRIOR FILING DATE: 2002-02-15
/ PRIOR APPLICATION NUMBER: US60/367,753
/ PRIOR FILING DATE: 2002-03-25
/ PRIOR APPLICATION NUMBER: US60/369,479
/ PRIOR FILING DATE: 2002-04-02
/ PRIOR APPLICATION NUMBER: US09/659,634
/ PRIOR FILING DATE: 2000-09-12
/ PRIOR APPLICATION NUMBER: US60/318,120
/ PRIOR FILING DATE: 2001-09-07
/ PRIOR APPLICATION NUMBER: US60/318,130
/ PRIOR FILING DATE: 2001-09-07
/ Remaining Prior Application data removed - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 794
/ SOFTWARE: Custom
/ SEQ ID NO 6
/ LENGTH: 311
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-236-392-6
```

```
Query Match      43.4%; Score 754.5; DB 15; Length 311;
Best Local Similarity 47.8%; Pred.No.9.3e-61;
Matches 143; Conservative 44; Mismatches 79; Indels 33; Gaps 4;

QY   20  FKNGPLPLOSLLQRLSRGCPCHLLLSLGLGLLLVILCVVGFQN-----SKFORDLVTLRT 74
Db   39  FLKPPPAQPLAQLRCSMVCFSLLLSFNILLVVICVTGSGEGHGRGAQLQLKSLKE 98

QY   75  DFSNFTSNTVAETQALTSQSSLEETIASLKAIEVEGFKQERQAGVSELOEHTTQKAHLGH 134
Db   99  AFSNFSSTLITEVQAISTHGSGVGDKITSLGAKLEKQQDLKA-----DHDALLFHLKH 152

QY   135  CPCHPSVCVPVHSEMLLRVQDLVQDLKKLTQVATLNNASTEGTCPCPVNWHQDSCYW 194
Db   153  FP-----VDLRFVACQMLLHNSGS-QRTCCPVNWHQSGCYW 190

QY   195  FSHSGMSWAEAEKYCOLKNAHLVINSREBNFQKYLGSAYTWMGLSDPEGAKWVDGT 254
Db   191  FSHSGKAWAEAEKYCOLENAHLVINSBEBQFIVQHTPFNTWIGLTDGDSWKWVDGT 250

QY   255  DYATGFQNKPKGPDQWQHGGLGGEDCAHFHPDGRWDDVCQRPYHWVCEAGLGQTSQ 313
Db   251  DYRHNKYNWAVTQPDNWHGHELGSEDCVEVQPDGRWDDFCIQVYRWVCEKERNATGE 309
```

Search completed: May 27, 2005, 12:52:11
Job time : 145 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 27, 2005, 12:44:45 ; Search time 22 Seconds
(without alignments)
1072.232 Million cell updates/sec

Title: US-10-829-107-4
Perfect score: 1738
Sequence: 1 MRYTFYFQYLENKVKVQGF.....QRPYHWCEAGLQTSQESH 316

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/prodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/prodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/prodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/prodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/prodata/1/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/prodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1738	100.0	316	3	US-09-111-470-4
2	1738	100.0	316	4	US-09-862-802A-4
3	1550.5	89.2	289	4	US-09-862-802A-13
4	1544	88.8	292	2	US-08-688-342-4
5	1544	88.8	292	2	US-09-113-788-4
6	1544	88.8	292	4	US-09-862-802A-12
7	1427.5	82.1	273	3	US-09-111-470-10
8	1427.5	82.1	273	4	US-09-862-802A-10
9	923.5	53.1	291	2	US-08-688-342-5
10	923.5	53.1	291	2	US-09-113-788-5
11	923.5	53.1	291	3	US-09-111-470-5
12	923.5	53.1	291	4	US-09-862-802A-5
13	923.5	53.1	311	4	US-09-949-016-9962
14	783	45.1	287	3	US-09-111-470-6
15	783	45.1	287	4	US-09-862-802A-6
16	745.5	42.9	272	4	US-09-949-016-9961
17	520	29.9	114	6	US14582-10
18	520	29.9	114	6	US14582-10
19	488	28.1	129	3	US-08-722-126A-10
20	488	28.1	129	5	PCT-US95-04258-10
21	454	26.1	114	6	US14582-8
22	454	26.1	114	6	US14582-8
23	432	24.9	114	6	US14582-14
24	432	24.9	114	6	US14582-14
25	410	23.6	114	6	US14582-13
26	410	23.6	114	6	US14582-13
27	347	20.0	77	4	US-09-531-056A-21

28 316 18.2 404 3 US-09-517-605-2 Sequence 2, Appli
29 287 16.5 327 1 US-08-365-103B-2 Sequence 2, Appli
30 283 16.3 287 1 US-08-365-103B-4 Sequence 4, Appli
31 283 16.3 300 1 US-08-365-103B-6 Sequence 6, Appli
32 279 16.1 292 4 US-09-535-521-2 Sequence 2, Appli
33 279 16.1 292 4 US-09-535-521-5 Sequence 5, Appli
34 278 16.0 208 4 US-09-535-521-20 Sequence 20, Appli
35 276.5 15.9 115 6 5514582-18 Patent No. 5514582
36 276.5 15.9 115 6 5514582-18 Patent No. 5514582
37 274.5 15.8 320 1 US-08-365-103B-10 Sequence 10, Appli
38 274.5 15.8 321 1 US-08-365-103B-8 Sequence 8, Appli
39 273 15.7 219 4 US-09-907-794A-377 Sequence 377, App
40 273 15.7 219 4 US-09-905-125A-377 Sequence 377, App
41 273 15.7 219 4 US-09-902-775A-377 Sequence 377, App
42 273 15.7 219 4 US-09-906-700A-377 Sequence 377, App
43 273 15.7 219 4 US-09-903-603A-377 Sequence 377, App
44 273 15.7 219 4 US-09-904-920A-377 Sequence 377, App
45 273 15.7 219 4 US-09-909-064-377 Sequence 377, App

ALIGNMENTS

RESULT 1
US-09-111-470-4
; Sequence 4, Application US/09111470
; Patent No. 6277959
; GENERAL INFORMATION:
; APPLICANT: Valladeau, Jenny
; APPLICANT: Ravel, Odile
; APPLICANT: Bates, Elizabeth E.M.
; APPLICANT: Ford, John
; APPLICANT: Saeland, Sem
; APPLICANT: Lebecque, Serge J.E.
; TITLE OF INVENTION: Mammalian Membrane Protein Genes;
; TITLE OF INVENTION: Related Reagents
; NUMBER OF SEQUENCES: 11

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DNAX Research Institute
; STREET: 901 California Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/111,470
; FILING DATE: 08-JUL-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/053,080
; FILING DATE: 09-JUL-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ching, Edwin P.
; REGISTRATION NUMBER: 34,090
; REFERENCE/DOCKET NUMBER: SF0695
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650)852-9196
; TELEFAX: (650)496-1200
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 316 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-111-470-4
Query Match 100.0%; Score 1738; DB 3; Length 316;
Best Local Similarity 100.0%; Pred. No. 1e-167;

Matches 316; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTRTYENFOYLENKVKVQGFKNKGPLQSLQRLRSGPCHLLLSLGLGLLLVLCVVG 60
DB 1 MTRTYENFOYLENKVKVQGFKNKGPLQSLQRLRSGPCHLLLSLGLGLLLVLCVVG 60
QY 61 QNSKFQDLVLTLDTSNFTSNFTVAEIQALTSQSSLEETIASLKAEEVGFQKQERQAGVS 120
DB 61 QNSKFQDLVLTLDTSNFTSNFTVAEIQALTSQSSLEETIASLKAEEVGFQKQERQAGVS 120
QY 121 ELQEHHTQKAHLGHCHPCPSVCVPVHSEMLLRVQQLVQDLKCLTCQVATLNNASTEGTC 180
DB 121 ELQEHHTQKAHLGHCHPCPSVCVPVHSEMLLRVQQLVQDLKCLTCQVATLNNASTEGTC 180
QY 181 CPVNWVEHODSCYFWSHSGMSWAEAEKYCOLKNAHLVWINSREONFVKYLGSAVTWVG 240
DB 181 CPVNWVEHODSCYFWSHSGMSWAEAEKYCOLKNAHLVWINSREONFVKYLGSAVTWVG 240
QY 241 LSDPEGAKWVDGTDYATGFQNWKPQDDWQGHGLGGGEDCAHFHPDGRWDDVCQRPY 300
DB 241 LSDPEGAKWVDGTDYATGFQNWKPQDDWQGHGLGGGEDCAHFHPDGRWDDVCQRPY 300
QY 301 HWVCEAGLGQTSQESH 316
DB 301 HWVCEAGLGQTSQESH 316

RESULT 2
US-09-862-802A-4
; Sequence 4, Application US/09862802A
; Patent No. 6756478
; GENERAL INFORMATION:
; APPLICANT: Valladeau, Jenny
; APPLICANT: Ravel, Odile
; APPLICANT: Bates, Elizabeth Ester Mary
; APPLICANT: Ford, John
; APPLICANT: Lebecque, Serge J.E.
; TITLE OF INVENTION: Isolated Mammalian Membrane Protein Genes; Related Reagents
; FILE REFERENCE: SF0695 B
; CURRENT APPLICATION NUMBER: US/09/862,802A
; CURRENT FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/053,080
; PRIOR FILING DATE: 1997-07-09
; PRIOR APPLICATION NUMBER: US 09/111,470
; PRIOR FILING DATE: 1998-07-08
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 4
; LENGTH: 316
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; FEATURE:
; OTHER INFORMATION: mammalian nucleic acid and protein
; NAME/KEY: misc feature
; LOCATION: (608)..(673)
; OTHER INFORMATION: short form lacks nucleotides 608-673
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (775)..(776)
; OTHER INFORMATION: ASGPRM (Table 2) has sequence insert encoding GEE between nucleot
; OTHER INFORMATION: ides 775-776
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1064)..(1064)
; OTHER INFORMATION: nucleotide 1064 of DCM2s may be A, which would encode Asn rather
; OTHER INFORMATION: than Asp at the residue numbered 270
US-09-862-802A-4

Query Match 100.0%; Score 1738; DB 4; Length 316;
Best Local Similarity 100.0%; Pred. No. 1e-167;
Matches 316; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTRTYENFOYLENKVKVQGFKNKGPLQSLQRLRSGPCHLLLSLGLGLLLVLCVVG 60
DB 1 MTRTYENFOYLENKVKVQGFKNKGPLQSLQRLRSGPCHLLLSLGLGLLLVLCVVG 60
QY 61 QNSKFQDLVLTLDTSNFTSNFTVAEIQALTSQSSLEETIASLKAEEVGFQKQERQAGVS 120
DB 61 QNSKFQDLVLTLDTSNFTSNFTVAEIQALTSQSSLEETIASLKAEEVGFQKQERQAGVS 120
QY 121 ELQEHHTQKAHLGHCHPCPSVCVPVHSEMLLRVQQLVQDLKCLTCQVATLNNASTEGTC 180
DB 121 ELQEHHTQKAHLGHCHPCPSVCVPVHSEMLLRVQQLVQDLKCLTCQVATLNNASTEGTC 180
QY 181 CPVNWVEHODSCYFWSHSGMSWAEAEKYCOLKNAHLVWINSREONFVKYLGSAVTWVG 240
DB 181 CPVNWVEHODSCYFWSHSGMSWAEAEKYCOLKNAHLVWINSREONFVKYLGSAVTWVG 240
QY 241 LSDPEGAKWVDGTDYATGFQNWKPQDDWQGHGLGGGEDCAHFHPDGRWDDVCQRPY 300
DB 241 LSDPEGAKWVDGTDYATGFQNWKPQDDWQGHGLGGGEDCAHFHPDGRWDDVCQRPY 300
QY 301 HWVCEAGLGQTSQESH 316
DB 301 HWVCEAGLGQTSQESH 316

RESULT 3
US-09-862-802A-13
; Sequence 13, Application US/09862802A
; Patent No. 6756478
; GENERAL INFORMATION:
; APPLICANT: Valladeau, Jenny
; APPLICANT: Ravel, Odile
; APPLICANT: Bates, Elizabeth Ester Mary
; APPLICANT: Ford, John
; APPLICANT: Lebecque, Serge J.E.
; TITLE OF INVENTION: Isolated Mammalian Membrane Protein Genes; Related Reagents
; FILE REFERENCE: SF0695 B
; CURRENT APPLICATION NUMBER: US/09/862,802A
; CURRENT FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/053,080
; PRIOR FILING DATE: 1997-07-09
; PRIOR APPLICATION NUMBER: US 09/111,470
; PRIOR FILING DATE: 1998-07-08
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 13
; LENGTH: 289
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: mammalian protein (DCMP2 short form)
US-09-862-802A-13

Query Match 89.2%; Score 1550.5; DB 4; Length 289;
Best Local Similarity 91.1%; Pred. No. 9.4e-149;
Matches 288; Conservative 1; Mismatches 0; Indels 27; Gaps 1;

QY 1 MTRTYENFOYLENKVKVQGFKNKGPLQSLQRLRSGPCHLLLSLGLGLLLVLCVVG 60
DB 1 MTRTYENFOYLENKVKVQGFKNKGPLQSLQRLRSGPCHLLLSLGLGLLLVLCVVG 60
QY 61 QNSKFQDLVLTLDTSNFTSNFTVAEIQALTSQSSLEETIASLKAEEVGFQKQERQAGVS 120
DB 61 QNSKFQDLVLTLDTSNFTSNFTVAEIQALTSQSSLEETIASLKAEEVGFQKQERQAGVS 117
QY 121 ELQEHHTQKAHLGHCHPCPSVCVPVHSEMLLRVQQLVQDLKCLTCQVATLNNASTEGTC 180
DB 118 -----VHSEMLLRVQQLVQDLKCLTCQVATLNNASTEGTC 153
QY 181 CPVNWVEHODSCYFWSHSGMSWAEAEKYCOLKNAHLVWINSREONFVKYLGSAVTWVG 240

Db 154 CPVNVVEHQDSCYWFSGMSWAEAEKYCOLKNAHLVWVINSREQNFVQKYLGSAYTMG 213
QY 241 LSPDEGAKWVDGTDYATGFGQNNKPGQPDWQGHGLGGGBDCAHFHDPDGRWDDVCQPY 300
Db 214 LSPDEGAKWVDGTDYATGFGQNNKPGQPDWQGHGLGGGBDCAHFHDPDGRWDDVCQPY 273
QY 301 HWVCEAGLGQTSQESH 316
Db 274 HWVCEAGLGQTSQESH 289

RESULT 4

US-08-688-342-4
; Sequence 4, Application US/08688342
; Patent No. 5871964
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Cocks, Benjamin G.
; APPLICANT: Goli, Surya K.
; APPLICANT: Hillman, Jennifer L.
; TITLE OF INVENTION: NOVEL HUMAN C-TYPE LECTIN
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/688,342
; FILING DATE: Filed Herewith
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0095-1 CIP
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 292 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 1235724
US-08-688-342-4

Query Match 88.8%; Score 1544; DB 2; Length 292;
Best Local Similarity 90.6%; Pred. No. 4.4e-148;
Matches 289; Conservative 0; Mismatches 0; Indels 30; Gaps 2;
QY 1 MTRTYENFQYLENKVKVQGFKNGLPLQSLQLRLRSGPCHLLLSGLGLLLLVICVVG 60
Db 1 MTRTYENFQYLENKVKVQGFKNGLPLQSLQLRLRSGPCHLLLSGLGLLLLVICVVG 60
QY 61 QNSKFQDLVTLRTDFTSNFTSNVVAETQALTSQSSLEETIASLKAEVEGFKQERQAGVS 120
Db 61 QNSKFQDLVTLRTDFTSNFTSNVVAETQALTSQSSLEETIASLKAEVEGFKQERQAGVS 117
QY 121 ELOEHTTQKAHLGHCHPCSPVCPVHSEMRLRQQLVQDLKCLTQCVATLNNN---ASTE 177
Db 118 -----VHSEMRLRQQLVQDLKCLTQCVATLNNNGEEASTE 153
QY 178 GTCCPVNVVEHQDSCYWFSGMSWAEAEKYCOLKNAHLVWVINSREQNFVQKYLGSAYT 237

Db 154 GTCCPVNVVEHQDSCYWFSGMSWAEAEKYCOLKNAHLVWVINSREQNFVQKYLGSAYT 213
QY 238 WMGLSDPEGAKWVDGTDYATGFGQNNKPGQPDWQGHGLGGGBDCAHFHDPDGRWDDVCQ 297
Db 214 WMGLSDPEGAKWVDGTDYATGFGQNNKPGQPDWQGHGLGGGBDCAHFHDPDGRWDDVCQ 273
QY 298 RPYHWVCEAGLGQTSQESH 316
Db 274 RPYHWVCEAGLGQTSQESH 292

RESULT 5

US-09-113-788-4
; Sequence 4, Application US/09113788
; Patent No. 5969104
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Cocks, Benjamin G.
; APPLICANT: Goli, Surya K.
; APPLICANT: Hillman, Jennifer L.
; TITLE OF INVENTION: NOVEL HUMAN C-TYPE LECTIN
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/113,788
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/688,342
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0095-1 CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 292 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 1235724
US-09-113-788-4

Query Match 88.8%; Score 1544; DB 2; Length 292;
Best Local Similarity 90.6%; Pred. No. 4.4e-148;
Matches 289; Conservative 0; Mismatches 0; Indels 30; Gaps 2;
QY 1 MTRTYENFQYLENKVKVQGFKNGLPLQSLQLRLRSGPCHLLLSGLGLLLLVICVVG 60
Db 1 MTRTYENFQYLENKVKVQGFKNGLPLQSLQLRLRSGPCHLLLSGLGLLLLVICVVG 60
QY 61 QNSKFQDLVTLRTDFTSNFTSNVVAETQALTSQSSLEETIASLKAEVEGFKQERQAGVS 120
Db 61 QNSKFQDLVTLRTDFTSNFTSNVVAETQALTSQSSLEETIASLKAEVEGFKQERQAGVS 117
QY 121 ELOEHTTQKAHLGHCHPCSPVCPVHSEMRLRQQLVQDLKCLTQCVATLNNN---ASTE 177

```
db 118 -----VHSEMLLRVQQLVQDLKKLTCCVATLNNNGEASTE 153
QY 178 GTCCPVNWEHODSCYFWSHSGMSWAEAEKYCOLKNAHLVWVINSREEQNFVKYLGSAVT 237
Db 154 GTCCPVNWEHODSCYFWSHSGMSWAEAEKYCOLKNAHLVWVINSREEQNFVKYLGSAVT 213
QY 238 WMGLSDPEGAKWVDGTDYATGFQNWKPQDDMQHGLGGGEDCAHFHPDGRWDDVQC 297
Db 214 WMGLSDPEGAKWVDGTDYATGFQNWKPQDDMQHGLGGGEDCAHFHPDGRWDDVQC 273
QY 298 RPYHWVCEAGLGQTSQESH 316
Db 274 RPYHWVCEAGLGQTSQESH 292

RESULT 6
US-09-862-802A-12
; Sequence 12, Application US/09862802A
; Patent No. 6756478
; GENERAL INFORMATION:
; APPLICANT: Valladeau, Jenny
; APPLICANT: Ravel, Odile
; APPLICANT: Bates, Elizabeth Ester Mary
; APPLICANT: Ford, John
; APPLICANT: Lebecque, Serge J.E.
; APPLICANT: Saeland, Sem
; TITLE OF INVENTION: Isolated Mammalian Membrane Protein Genes; Related Reagents
; FILE REFERENCE: SF0695 B
; CURRENT APPLICATION NUMBER: US/09/862,802A
; CURRENT FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/053,080
; PRIOR FILING DATE: 1997-07-09
; PRIOR APPLICATION NUMBER: US 09/111,470
; PRIOR FILING DATE: 1998-07-08
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 12
; LENGTH: 292
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: mammalian protein (ASGPRm is a macrophage derived ASGPR)
US-09-862-802A-12

Query Match 88.8%; Score 1544; DB 4; Length 292;
Best Local Similarity 90.6%; Pred. No. 4.4e-148;
Matches 289; Conservative 0; Mismatches 0; Indels 30; Gaps 2;

QY 1 MTRTYENFOYLENKVKVQGFKNQPLQSLQRLRSRGPCHLLSLGLGLLLLVIIICVVG 60
Db 1 MTRTYENFOYLENKVKVQGFKNQPLQSLQRLRSRGPCHLLSLGLGLLLLVIIICVVG 60
QY 61 QNSKFORDLVLTDFSNFTSNFTVAEIQALTSQSSLEETIASLKAEEVGFQKQERQAGVS 120
Db 61 QNSKFORDLVLTDFSNFTSNFTVAEIQALTSQSSLEETIASLKAEEVGFQKQERQA--- 117
QY 121 ELQEHHTTQKAHLGHCHPCPSVCVPVHSEMLLRVQQLVQDLKKLTCCVATLNNN---ASTE 177
Db 118 -----VHSEMLLRVQQLVQDLKKLTCCVATLNNNGEASTE 153
QY 178 GTCCPVNWEHODSCYFWSHSGMSWAEAEKYCOLKNAHLVWVINSREEQNFVKYLGSAVT 237
Db 154 GTCCPVNWEHODSCYFWSHSGMSWAEAEKYCOLKNAHLVWVINSREEQNFVKYLGSAVT 213
QY 238 WMGLSDPEGAKWVDGTDYATGFQNWKPQDDMQHGLGGGEDCAHFHPDGRWDDVQC 297
Db 214 WMGLSDPEGAKWVDGTDYATGFQNWKPQDDMQHGLGGGEDCAHFHPDGRWDDVQC 273
QY 298 RPYHWVCEAGLGQTSQESH 316
Db 274 RPYHWVCEAGLGQTSQESH 292
```

```
RESULT 7
US-09-111-470-10
; Sequence 10, Application US/09111470
; Patent No. 6277959
; GENERAL INFORMATION:
; APPLICANT: Valladeau, Jenny
; APPLICANT: Ravel, Odile
; APPLICANT: Bates, Elizabeth E.M.
; APPLICANT: Ford, John
; APPLICANT: Saeland, Sem
; APPLICANT: Lebecque, Serge J.E.
; TITLE OF INVENTION: Mammalian Membrane Protein Genes;
; TITLE OF INVENTION: Related Reagents
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DNAX Research Institute
; STREET: 901 California Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/111,470
; FILING DATE: 08-JUL-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/053,080
; FILING DATE: 09-JUL-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ching, Edwin P.
; REGISTRATION NUMBER: 34,090
; REFERENCE/DOCKET NUMBER: SF0695
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650)852-9196
; TELEFAX: (650)498-1200
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 273 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-111-470-10
```

```
Query Match 82.1%; Score 1427.5; DB 3; Length 273;
Best Local Similarity 84.6%; Pred. No. 2.6e-136;
Matches 270; Conservative 0; Mismatches 0; Indels 49; Gaps 3;

QY 1 MTRTYENFOYLENKVKVQGFKNQPLQSLQRLRSRGPCHLLSLGLGLLLLVIIICVVG 60
Db 1 MTRTYENFOYLENKVKVQGFKNQPLQSLQRLRSRGPCHLLSLGLGLLLLVIIICVVG 41
QY 61 QNSKFORDLVLTDFSNFTSNFTVAEIQALTSQSSLEETIASLKAEEVGFQKQERQAGVS 120
Db 42 QNSKFORDLVLTDFSNFTSNFTVAEIQALTSQSSLEETIASLKAEEVGFQKQERQA--- 98
QY 121 ELQEHHTTQKAHLGHCHPCPSVCVPVHSEMLLRVQQLVQDLKKLTCCVATLNNN---ASTE 177
Db 99 -----VHSEMLLRVQQLVQDLKKLTCCVATLNNNGEASTE 134
QY 178 GTCCPVNWEHODSCYFWSHSGMSWAEAEKYCOLKNAHLVWVINSREEQNFVKYLGSAVT 237
Db 135 GTCCPVNWEHODSCYFWSHSGMSWAEAEKYCOLKNAHLVWVINSREEQNFVKYLGSAVT 194
QY 238 WMGLSDPEGAKWVDGTDYATGFQNWKPQDDMQHGLGGGEDCAHFHPDGRWDDVQC 297
Db 195 WMGLSDPEGAKWVDGTDYATGFQNWKPQDDMQHGLGGGEDCAHFHPDGRWDDVQC 254
```

QY 298 RPYHWVCEAGLGQTSQESH 316
Db 255 RPYHWVCEAGLGQTSQESH 273

RESULT 8

US-09-862-802A-10
; Sequence 10, Application US/09862802A
; Patent No. 6756478
; GENERAL INFORMATION:
; APPLICANT: Valladeau, Jenny
; APPLICANT: Ravel, Odile
; APPLICANT: Bates, Elizabeth Ester Mary
; APPLICANT: Ford, John
; APPLICANT: Lebecque, Serge J.B.
; APPLICANT: Saeland, Sem
; TITLE OF INVENTION: Isolated Mammalian Membrane Protein Genes; Related Reagents
; FILE REFERENCE: SF0695 B
; CURRENT APPLICATION NUMBER: US/09/862,802A
; CURRENT FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/053,080
; PRIOR FILING DATE: 1997-07-09
; PRIOR APPLICATION NUMBER: US 09/111,470
; PRIOR FILING DATE: 1998-07-08
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 273
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: mammalian nucleic acid and protein
US-09-862-802A-10

Query Match 82.1%; Score 1427.5; DB 4; Length 273;
Best Local Similarity 84.6%; Pred. No. 2.6e-136;
Matches 270; Conservative 0; Mismatches 0; Indels 49; Gaps 3;
QY 1 MTRTYENFOYLENKVKVQGFKNGLPLQSLQLRLSGPCHLLSLGLGLLLVLCVVG 60
Db 1 MTRTYENFOYLENKVKVQGFKNGLPLQSLQLRLSGPCHLLSLGLGLLLVLCVVG 41
QY 61 QNSKFORDLVLTDFSNFTNTVAEIQALTSQSSLEETIASLKAEEVGFQERQGV 120
Db 42 QNSKFORDLVLTDFSNFTNTVAEIQALTSQSSLEETIASLKAEEVGFQERQGV 98
QY 121 ELQEHHTTQKAHLGHCPCSPVCPVHSEMLLRVQQLVQDLKLTCCQVATLNNN 177
Db 99 -----VHSEMLLRVQQLVQDLKLTCCQVATLNNN 134
QY 178 GTCCPVNVVHSDSCYWFSGMSWAEAEKYCOLKNAHLVIVNSREEQNFVQKYLGSAYT 237
Db 135 GTCCPVNVVHSDSCYWFSGMSWAEAEKYCOLKNAHLVIVNSREEQNFVQKYLGSAYT 194
QY 238 WMGLSDPEGAKWVDGTDYATGFQNKWPGQDDWQHGGLGGEDCAHFHPDGRWDDVCQ 297
Db 195 WMGLSDPEGAKWVDGTDYATGFQNKWPGQDDWQHGGLGGEDCAHFHPDGRWDDVCQ 254
QY 298 RPYHWVCEAGLGQTSQESH 316
Db 255 RPYHWVCEAGLGQTSQESH 273

RESULT 9

US-08-688-342-5
; Sequence 5, Application US/08688342
; Patent No. 5871964
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Cocks, Benjamin G.
; APPLICANT: Goli, Surya K.
; APPLICANT: Hillman, Jennifer L.
; TITLE OF INVENTION: NOVEL HUMAN C-TYPE LECTIN

; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/688,342
; FILING DATE: Filed Herewith
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0095-1 CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 291 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 179079
US-08-688-342-5

Query Match 53.1%; Score 923.5; DB 2; Length 291;
Best Local Similarity 54.6%; Pred. No. 3.5e-85;
Matches 172; Conservative 47; Mismatches 67; Indels 29; Gaps 5;
QY 1 MTRTYENFOYLENKVK-VQGFKNGLPLQSLQLRLSGPCHLLSLGLGLLLVLCVVG 59
Db 1 MTRTYENFOYLENKVKVQGFKNGLPLQSLQLRLSGPCHLLSLGLGLLLVLCVVG 60
QY 60 FQNSKFORDLVLTDFSNFTNTVAEIQALTSQSSLEETIASLKAEEVGFQERQGV 119
Db 61 SQNSQLQEELRGLRFTSNFTASTAQVGLSTQGGNVGRKMKLSQLE--KQK--- 114
QY 120 SELQEHHTTQKAHLGHCPCSPVCPVHSEMLLRVQQLVQDLKLTCCQVATLNNN 179
Db 115 -DUSED-----HSLLHLVKQFVSDRLSLSCQMAALQNGS-ERT 152
QY 180 CCPNVNVHSDSCYWFSGMSWAEAEKYCOLKNAHLVIVNSREEQNFVQKYLGSAYTM 239
Db 153 CCPNVNVHSDSCYWFSGMSWAEAEKYCOLKNAHLVIVNSREEQNFVQKYLGSAYTM 212
QY 240 GLSDPEGAKWVDGTDYATGFQNKWPGQDDWQHGGLGGEDCAHFHPDGRWDDVCQ 299
Db 213 GLHDGPNKWDGTDYATGFQNKWPGQDDWQHGGLGGEDCAHFHPDGRWDDVCQ 272
QY 300 YHWVCEAGLGQTSQESH 314
Db 273 YRWVCEAGLGQTSQESH 287

RESULT 10

US-09-113-788-5
; Sequence 5, Application US/09113788
; Patent No. 5969104
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Cocks, Benjamin G.
; APPLICANT: Goli, Surya K.
; APPLICANT: Hillman, Jennifer L.


```
RESULT 12
US-09-862-802A-5
; Sequence 5, Application US/09862802A
; Patent No. 6756478
; GENERAL INFORMATION:
; APPLICANT: Valladeau, Jenny
; APPLICANT: Ravel, Odile
; APPLICANT: Bates, Elizabeth Ester Mary
; APPLICANT: Ford, John
; APPLICANT: Lebecque, Serge J.E.
; APPLICANT: Saeland, Sem
; TITLE OF INVENTION: Isolated Mammalian Membrane Protein Genes; Related Reagents
; FILE REFERENCE: SF0695 B
; CURRENT APPLICATION NUMBER: US/09/862,802A
; CURRENT FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/053,080
; PRIOR FILING DATE: 1997-07-09
; PRIOR APPLICATION NUMBER: US 09/111,470
; PRIOR FILING DATE: 1998-07-08
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 291
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: mammalian protein
US-09-862-802A-5

Query Match 53.1%; Score 923.5; DB 4; Length 291;
Best Local Similarity 54.6%; Pred. No. 3.5e-85;
Matches 172; Conservative 47; Mismatches 67; Indels 29; Gaps 5;

QY 1 MTRTYENFOYLENKVK-VQGFKNGLPLQSLQLRLSGPCCHLLSLGLGLLLVLCVVG 59
Db 1 MTKEYQDLQHLNDESDHQLRKGPPPPQPLLQRLCSGPRLLLSLGLSLLLVVVCVIG 60
QY 60 FQNSKFQDRLVTLRTDFSNFTNTVAEIQALTSQSSLEETIASLKAEEVGFQKQERQAGV 119
Db 61 SQNSQLQEELRLGRLRETFNSTASTAQVKGSLTQGGNVGRKMSLSQLE--KQK- 114
QY 120 SELQHTTQKAHLGHCPCFSCVCPVHSEMLLRVQQLVQDLKLTCCQVATLNNASTEGT 179
Db 115 -DLS- 179
QY 180 CCPVNWVHODSCYFWSHSGMSWAEAEKYCOLKNAHLVWINSREEQNFVQKYLGSAYTWM 239
Db 153 CCPVNWVHERSCYFWSRSGKAWADADNYCRLEDAHLVVVTSWEEQKFVQHHIGPVNTWM 212
QY 240 GLSDPEGAKWVDGTDYATGQFNWKPQDDWQGHGLGGEDCAHFHPDGRWDDVCQRP 299
Db 213 GLHDQNGPWKWDGTDYETGFKNRPQDDWYGHGLGGEDCAHFHTDGRWDDVCQRP 272
QY 300 YHWVCEAGLGQTSQE 314
Db 273 YRWVCE TELDKASQE 287

RESULT 13
US-09-949-016-9962
; Sequence 9962, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C0001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
```

```
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9962
; LENGTH: 311
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-9962

Query Match 53.1%; Score 923.5; DB 4; Length 311;
Best Local Similarity 54.6%; Pred. No. 3.9e-85;
Matches 172; Conservative 47; Mismatches 67; Indels 29; Gaps 5;

QY 1 MTRTYENFOYLENKVK-VQGFKNGLPLQSLQLRLSGPCCHLLSLGLGLLLVLCVVG 59
Db 21 MTKEYQDLQHLNDESDHQLRKGPPPPQPLLQRLCSGPRLLLSLGLSLLLVVVCVIG 80
QY 60 FQNSKFQDRLVTLRTDFSNFTNTVAEIQALTSQSSLEETIASLKAEEVGFQKQERQAGV 119
Db 81 SQNSQLQEELRLGRLRETFNSTASTAQVKGSLTQGGNVGRKMSLSQLE--KQK- 134
QY 120 SELQHTTQKAHLGHCPCFSCVCPVHSEMLLRVQQLVQDLKLTCCQVATLNNASTEGT 179
Db 135 -DLS- 179
QY 180 CCPVNWVHODSCYFWSHSGMSWAEAEKYCOLKNAHLVWINSREEQNFVQKYLGSAYTWM 239
Db 173 CCPVNWVHERSCYFWSRSGKAWADADNYCRLEDAHLVVVTSWEEQKFVQHHIGPVNTWM 232
QY 240 GLSDPEGAKWVDGTDYATGQFNWKPQDDWQGHGLGGEDCAHFHPDGRWDDVCQRP 299
Db 233 GLHDQNGPWKWDGTDYETGFKNRPQDDWYGHGLGGEDCAHFHTDGRWDDVCQRP 292
QY 300 YHWVCEAGLGQTSQE 314
Db 293 YRWVCE TELDKASQE 307

RESULT 14
US-09-111-470-6
; Sequence 6, Application US/09111470
; Patent No. 6277959
; GENERAL INFORMATION:
; APPLICANT: Valladeau, Jenny
; APPLICANT: Ravel, Odile
; APPLICANT: Bates, Elizabeth E.M.
; APPLICANT: Ford, John
; APPLICANT: Saeland, Sem
; APPLICANT: Lebecque, Serge J.E.
; TITLE OF INVENTION: Mammalian Membrane Protein Genes;
; TITLE OF INVENTION: Related Reagents
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DNAX Research Institute
; STREET: 901 California Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/111,470
; FILING DATE: 08-JUL-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/053,080
; FILING DATE: 09-JUL-1997
; ATTORNEY/AGENT INFORMATION:
```

```
/ NAME: Ching, Edwin P.
/ REGISTRATION NUMBER: 34,090
/ REFERENCE/DOCKET NUMBER: SF0695
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (650)852-9196
/ TELEFAX: (650)496-1200
/ INFORMATION FOR SEQ ID NO: 6:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 287 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: not relevant
/ TOPOLOGY: linear
/ MOLECULE TYPE: peptide
US-09-111-470-6

Query Match 45.1%; Score 783; DB 3; Length 287;
Best Local Similarity 46.6%; Pred. No. 6.1e-71;
Matches 146; Conservative 50; Mismatches 89; Indels 28; Gaps 3;

QY 1 MTRTYENFOYLENKVKVQGFKNGLPLQSLQLRLRSGPCCHLLLSLGLGLLLVLCVVG 60
DB 1 MAKDFODIQQLSSEENDHPFHQGPAPLAQLRLCSWVCFSLLALSNILLVVICVTGS 60
QY 61 QNSKFORDLVTLRTDPSNFTSNTVAIEIOALTSQSSLEETIASLKAEVGFQKQROAGVS 120
DB 61 QSAQLQAEIQLSKEAFSNFSSSTLTVEVQAIETHGGSGVDKITSLGAKLEKQQQDLKA--- 117
QY 121 ELQEHHTQKAHLGHCHPCPSVCVPVHSEMLLRVQQLVQDLKLTQCVATLNNASTEGTC 180
DB 118 ---DHDALLFHLKHFP-----VDLRFVACQMELLHSNGS-QRTC 152
QY 181 CPVNVVEHQDSCYWFSGHSGMSWAEAEKYCOLKNAHLVWINSREEQNFVKYLGSAYTWMG 240
DB 153 CPVNVVEHQDSCYWFSGHSGKAWAEAEKYCOLENAHLVWINSWEEQKFIVQHTNPNFTWIG 212
QY 241 LSDPEGANKWVDGTDYATGFQNWKPQDPDQGHGLGGGEDCAHFHPDGRWDDVVCORPY 300
DB 213 LTDSGSKWVGDGTDYRHNKYNWAVTQPDNWHGHELGSSEDCVEVQPDGRWDDDFCLQVY 272
QY 301 HWVCEAGLGQTSQ 313
DB 273 RNVCEKRRNATGE 285

RESULT 15
US-09-862-802A-6
/ Sequence 6, Application US/09862802A
/ Patent No. 6756478
/ GENERAL INFORMATION:
/ APPLICANT: Valladeau, Jenny
/ APPLICANT: Ravel, Odile
/ APPLICANT: Bates, Elizabeth Ester Mary
/ APPLICANT: Ford, John
/ APPLICANT: Lebecque, Serge J.E.
/ APPLICANT: Saeland, Sem
/ TITLE OF INVENTION: Isolated Mammalian Membrane Protein Genes; Related Reagents
/ FILE REFERENCE: SF0695 B
/ CURRENT APPLICATION NUMBER: US/09/862,802A
/ CURRENT FILING DATE: 2001-05-22
/ PRIOR APPLICATION NUMBER: US 60/053,080
/ PRIOR FILING DATE: 1997-07-09
/ PRIOR APPLICATION NUMBER: US 09/111,470
/ PRIOR FILING DATE: 1998-07-08
/ NUMBER OF SEQ ID NOS: 13
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 6
/ LENGTH: 287
/ TYPE: PRT
/ ORGANISM: Unknown
/ FEATURE:
/ OTHER INFORMATION: mammalian protein
US-09-862-802A-6
```

```
Query Match 45.1%; Score 783; DB 4; Length 287;
Best Local Similarity 46.6%; Pred. No. 6.1e-71;
Matches 146; Conservative 50; Mismatches 89; Indels 28; Gaps 3;

QY 1 MTRTYENFOYLENKVKVQGFKNGLPLQSLQLRLRSGPCCHLLLSLGLGLLLVLCVVG 60
DB 1 MAKDFODIQQLSSEENDHPFHQGPAPLAQLRLCSWVCFSLLALSNILLVVICVTGS 60
QY 61 QNSKFORDLVTLRTDPSNFTSNTVAIEIOALTSQSSLEETIASLKAEVGFQKQROAGVS 120
DB 61 QSAQLQAEIQLSKEAFSNFSSSTLTVEVQAIETHGGSGVDKITSLGAKLEKQQQDLKA--- 117
QY 121 ELQEHHTQKAHLGHCHPCPSVCVPVHSEMLLRVQQLVQDLKLTQCVATLNNASTEGTC 180
DB 118 ---DHDALLFHLKHFP-----VDLRFVACQMELLHSNGS-QRTC 152
QY 181 CPVNVVEHQDSCYWFSGHSGMSWAEAEKYCOLKNAHLVWINSREEQNFVKYLGSAYTWMG 240
DB 153 CPVNVVEHQDSCYWFSGHSGKAWAEAEKYCOLENAHLVWINSWEEQKFIVQHTNPNFTWIG 212
QY 241 LSDPEGANKWVDGTDYATGFQNWKPQDPDQGHGLGGGEDCAHFHPDGRWDDVVCORPY 300
DB 213 LTDSGSKWVGDGTDYRHNKYNWAVTQPDNWHGHELGSSEDCVEVQPDGRWDDDFCLQVY 272
QY 301 HWVCEAGLGQTSQ 313
DB 273 RNVCEKRRNATGE 285
```

Search completed: May 27, 2005, 12:49:19
Job time : 24 secs

Result No.	Query #			DB	ID	Description
	Score	Match	Length			
1	1738	100.0	316	2	AAW88125	AAw88125 Primate D
2	1718.5	98.9	319	5	ABG66703	ABg66703 Human nov
3	1718.5	98.9	319	6	ABu69170	ABu69170 Human NOV
4	1651.5	95.0	315	8	ADO08343	Ado08343 Human NOV
5	1427.5	82.1	273	2	AAW88129	AAw88129 Variant p
6	1410.5	81.2	267	6	ABU69169	ABu69169 Human NOV
7	1410.5	81.2	267	8	ADO08341	Ado08341 Human NOV
8	923.5	53.1	291	2	AAW15245	AAw15245 Astialogy
9	908.5	52.3	319	8	ABW82572	ABm82572 Human dia
10	908.5	52.3	319	8	ABW82574	ABm82574 Human dia
11	902	51.9	314	8	ABW82573	ABm82573 Human dia
12	902	51.9	314	8	ABW82575	ABm82575 Human dia
13	862	49.6	290	8	ABW82571	ABm82571 Human dia
14	862	49.6	290	8	ABW82570	ABm82570 Human dia
15	838	48.2	274	2	AAW15250	AAw15250 Astialogy
16	783	45.1	287	2	AAW15246	AAw15246 Astialogy
17	783	45.1	287	6	ABO01292	ABo01292 Human pro
18	778	44.8	232	2	AAW15249	AAw15249 Astialogy
19	776	44.6	252	8	ABW82576	ABm82576 Human dia
20	770.5	44.3	292	8	ADP23959	ADp23959 PRO poly
21	768	44.2	287	8	ADN95956	ADn95956 Human NOV
22	760	43.7	306	6	ABO01291	ABo01291 Human pro
23	760	43.7	306	8	ADN95954	ADn95954 Human NOV
24	754.5	43.4	311	6	ABO01290	ABo01290 Human pro
25	754.5	43.4	311	7	ADe63359	ADe63359 Human pro

SQ	Sequence 316 AA;	
	Query Match	100.0%; Score 1738; DB 2; Length 316;
	Best Local Similarity	100.0%; Pred. No. 1.9e-154;
	Matches 316; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	1 MTRTYENFOYLENKVKVQGFKNGLPLQSLQLRLRSGPCHLLLSLGLGLLLVIIICVVG 60	
DB	1 MTRTYENFOYLENKVKVQGFKNGLPLQSLQLRLRSGPCHLLLSLGLGLLLVIIICVVG 60	
QY	61 QNSKFORDLVLTIRDFSNFTSNTVAEIQALTSQSSSLEETIASLKAEEVGFQKQERQAGVS 120	
DB	61 QNSKFORDLVLTIRDFSNFTSNTVAEIQALTSQSSSLEETIASLKAEEVGFQKQERQAGVS 120	
QY	121 ELQEHHTTQKAHLGHCHPCPSVCVPVHSEMLLRVQQLVQDLKKLTQCVATLNNASTEGTC 180	
DB	121 ELQEHHTTQKAHLGHCHPCPSVCVPVHSEMLLRVQQLVQDLKKLTQCVATLNNASTEGTC 180	
QY	181 CPVNWVEHQDSCYWFSGHSGMSWAEAEKYCOLKNAHLVWINSREEQNFVKYLGSAVTWNG 240	
DB	181 CPVNWVEHQDSCYWFSGHSGMSWAEAEKYCOLKNAHLVWINSREEQNFVKYLGSAVTWNG 240	
QY	241 LSDPEGAKWVDGTDYATGFQNWKPQDDWQHGHLGGEDCAHFPHDPGRWDDVVCQPY 300	
DB	241 LSDPEGAKWVDGTDYATGFQNWKPQDDWQHGHLGGEDCAHFPHDPGRWDDVVCQPY 300	
QY	301 HWCEAGLGQTSQESH 316	
DB	301 HWCEAGLGQTSQESH 316	
RESULT 2		
ID	ABG66703 standard; protein; 319 AA.	
XX	ABG66703;	
AC		
DT	30-AUG-2002 (first entry)	
XX	Human novel polypeptide #38.	
DE		
XX	Human; inflammatory condition; shock; sepsis; immune response; cancer;	
KW	wound healing; central nervous system disease; haematopoesis;	
KW	peripheral nervous system disease; amyotrophic lateral sclerosis; tendon;	
KW	myeloid cell disorder; lymphoid cell disorder; platelet disorder; bone;	
KW	cartilage; ligament; nerve tissue; ulcer; osteoporosis; osteoarthritis;	
KW	bone degenerative disorder; periodontal disease; reperfusion injury;	
KW	lung fibrosis; liver fibrosis; autoimmune disease; bacterial infection;	
KW	allergic condition; thrombolysis; thrombosis; coagulation disorder;	
XX	fungal infection.	
OS	Homo sapiens.	
XX		
PN	WO200244340-A2.	
XX		
PD	06-JUN-2002.	
XX		
PF	30-NOV-2001; 2001WO-US047004.	
XX		
PR	30-NOV-2000; 2000US-00728952.	
XX		
PA	(HYSE-) HYSEQ INC.	
XX		
PI	Tang YT, Goodrich RW, Liu C, Zhou P, Asundi V, Wang J, Wang D;	
PI	Yamazaki V, Ujwal ML, Drmanac RT;	
XX		
DR	WPI; 2002-508509/54.	
DR	N-PSDB; ABK94927.	
XX		
XX	Novel nucleic acids and polypeptides for diagnosis, treatment of	
PT	inflammatory, autoimmune, nervous system, myeloid or lymphoid cell	
PT	disorders, cancer and promoting wound healing.	
XX		

PS	Claim 10; Page 607-608; 672pp; English.	
XX		
CC	The invention relates to human novel polynucleotides and associated	
CC	polypeptides. The polynucleotides and polypeptides are useful for	
CC	treating inflammatory conditions such as arthritis, nephritis, Crohn's	
CC	disease, ischaemia-reperfusion injury, shock, sepsis, immune responses	
CC	and cancer and for promoting wound healing. The sequences are used to	
CC	induce the proliferation of neural cells and regeneration of nerve and	
CC	brain tissue, and are useful for the treatment of central and peripheral	
CC	nervous system diseases and neuropathies, such as Alzheimer's disease,	
CC	Parkinson's disease, Huntington's disease and amyotrophic lateral	
CC	sclerosis. The sequences are involved in chemotactic or chemokinetic	
CC	activity, regulation of haematopoesis, treatment of myeloid or lymphoid	
CC	cell disorders and platelet disorders such as thrombocytopenia,	
CC	regeneration of bone, cartilage, tendon, ligament and/or nerve tissue	
CC	growth, tissue repair, healing of burns, incisions, ulcers, treatment of	
CC	osteoporosis, osteoarthritis, bone degenerative disorders and periodontal	
CC	disease. The sequences of the invention are also useful for gut	
CC	protection or regeneration and treatment of lung or liver fibrosis,	
CC	reperfusion injury in various tissues, immune deficiencies and disorders	
CC	including severe combined immunodeficiency (SCID), bacterial or fungal	
CC	infections, autoimmune disorders e.g. multiple sclerosis and myasthenia	
CC	gravis, allergic conditions such as asthma, thrombolysis or thrombosis	
CC	and coagulation disorders. Sequences ABG66666-ABG66758 represent human	
CC	novel polypeptides of the invention	
XX		
SQ	Sequence 319 AA;	
	Query Match	98.9%; Score 1718.5; DB 5; Length 319;
	Best Local Similarity	98.7%; Pred. No. 1.3e-152;
	Matches 315; Conservative	0; Mismatches 1; Indels 3; Gaps 1;
QY	1 MTRTYENFOYLENKVKVQGFKNGLPLQSLQLRLRSGPCHLLLSLGLGLLLVIIICVVG 60	
DB	1 MTRTYENFOYLENKVKVQGFKNGLPLQSLQLRLRSGPCHLLLSLGLGLLLVIIICVVG 60	
QY	61 QNSKFORDLVLTIRDFSNFTSNTVAEIQALTSQSSSLEETIASLKAEEVGFQKQERQAGVS 120	
DB	61 QNSKFORDLVLTIRDFSNFTSNTVAEIQALTSQSSSLEETIASLKAEEVGFQKQERQAGVS 120	
QY	121 ELQEHHTTQKAHLGHCHPCPSVCVPVHSEMLLRVQQLVQDLKKLTQCVATLNN---ASTE 177	
DB	121 ELQEHHTTQKAHLGHCHPCPSVCVPVHSEMLLRVQQLVQDLKKLTQCVATLNNNGEASTE 180	
QY	178 GTCCPVNVEHQDSCYWFSGHSGMSWAEAEKYCOLKNAHLVWINSREEQNFVKYLGSAVT 237	
DB	181 GTCCPVNVEHQDSCYWFSGHSGMSWAEAEKYCOLKNAHLVWINSREEQNFVKYLGSAVT 240	
QY	238 WMLSDPEGAKWVDGTDYATGFQNWKPQDDWQHGHLGGEDCAHFPHDPGRWDDVVCQ 297	
DB	241 WMLSDPEGAKWVDGTDYATGFQNWKPQDDWQHGHLGGEDCAHFPHDPGRWDDVVCQ 300	
QY	298 RPYHWVCEAGLGQTSQESH 316	
DB	301 RPYHWVCEAGLGQTSQESH 319	
RESULT 3		
ID	ABU69170 standard; protein; 319 AA.	
XX	ABU69170;	
AC		
DT	02-JUN-2003 (first entry)	
XX	Human NOVX polypeptide #45.	
DE		
XX	Human; NOVX; metabolic disorder; diabetes; infectious disease; obesity;	
KW	anorexia; cancer; cardiovascular disorder; asthma; neurogenesis;	
KW	neurodegenerative disorder; epilepsy; immune disorder; osteoarthritis;	
KW	haematopoietic disorder; inflammatory skin disorder; dyslipidemia;	
KW	haematopoesis; wound healing; angiogenesis; bacterial infection;	
KW	viral infection; fungal infection; helminthic infection; atherosclerosis;	

KW protozoal infection; hypertension.
 XX Homo sapiens.
 OS
 XX
 XX WQ200290504-A2.
 XX
 XX
 PD 14-NOV-2002.
 XX
 XX 02-MAY-2002; 2002WO-US014342.
 PF
 XX
 XX 03-MAY-2001; 2001US-0288395P.
 PR 04-MAY-2001; 2001US-0288900P.
 PR 07-MAY-2001; 2001US-0289087P.
 PR 14-MAY-2001; 2001US-0290753P.
 PR 15-MAY-2001; 2001US-0291189P.
 PR 16-MAY-2001; 2001US-0291243P.
 PR 18-MAY-2001; 2001US-0292001P.
 PR 21-MAY-2001; 2001US-0292374P.
 PR 22-MAY-2001; 2001US-0292587P.
 PR 23-MAY-2001; 2001US-0293107P.
 PR 29-MAY-2001; 2001US-0294110P.
 PR 30-MAY-2001; 2001US-0294434P.
 PR 31-MAY-2001; 2001US-0294827P.
 PR 18-JUN-2001; 2001US-0298988P.
 PR 31-JUL-2001; 2001US-0308901P.
 PR 17-AUG-2001; 2001US-0313388P.
 PR 21-AUG-2001; 2001US-0313851P.
 PR 21-AUG-2001; 2001US-0313937P.
 PR 17-SEP-2001; 2001US-0322701P.
 PR 17-SEP-2001; 2001US-0322802P.
 PR 25-SEP-2001; 2001US-0324757P.
 PR 27-SEP-2001; 2001US-0325314P.
 PR 27-SEP-2001; 2001US-0325682P.
 PR 21-NOV-2001; 2001US-0332129P.
 PR 03-DEC-2001; 2001US-0336882P.
 PR 14-DEC-2001; 2001US-0340305P.
 PR 01-MAY-2002; 2002US-00138588.
 XX
 XX (CURA-) CURAGEN CORP.
 XX
 XX Alsobrook JP, Anderson DW, Boldog FL, Burgess CE, Casman SJ;
 PI Chapoval A, Edinger S, Gerlach V, Gorman L, Gunther E, Guo X;
 PI Kekuda R, Lepley DM, Li L, Liu X, Malyanar UM, Miller CE;
 PI Millet I, Padigaru M, Patturajan M, Pena CEA, Rieger DK, Shenoy SG;
 PI Shimkets RA, Spytek KA, Taupier RJ, Vernet CNM, Voss EZ;
 PI Zerhusen BD;
 XX
 DR WPI; 2003-103512/09.
 DR N-PSDB; ACA10155.
 XX
 XX New isolated NOVX polypeptides and polynucleotides, useful for
 PT preventing, diagnosing or treating NOVX-associated disorders, e.g.
 PT osteoarthritis, obesity, atherosclerosis, cancer, Parkinson's disease,
 PT asthma, or infections.
 XX
 XX Claim 2; Page 170; 340pp; English.
 PS
 XX
 XX The invention relates to human NOVX polypeptides and the polynucleotides
 CC encoding them. The polypeptides, polynucleotides and antibodies that bind
 CC immunospecifically to the polypeptides are useful in the manufacture of a
 CC medicament for treating a syndrome associated with a human disease,
 CC preferably a NOVX-associated disorder. The sequences are useful for
 CC treating, preventing or diagnosing diseases such as metabolic disorders,
 CC diabetes, obesity, infectious diseases (viral, bacterial, fungal,
 CC helminthic, and protozoal), anorexia, cancer, cardiovascular disorders
 CC (e.g. hypertension, atherosclerosis), neurodegenerative disorders (e.g.
 CC Alzheimer's disease, Parkinson's disease), epilepsy, immune disorders,
 CC osteoarthritis, haematopoietic disorders, inflammatory skin disorders,
 CC asthma and various dyslipidemias. The nucleic acids and polypeptides may
 CC also be used as targets for the identification of small molecules that
 CC modulate or inhibit e.g. neurogenesis, cell differentiation, cell
 CC proliferation, haematopoiesis, wound healing and angiogenesis, and in the
 CC generation of antibodies that bind immunospecifically to NOVX substances

CC for use in therapeutic or diagnostic methods. The nucleic acids are
 CC further used as hybridisation probes, and in chromosome mapping, tissue
 CC typing, preventive medicine and pharmacogenomics. Sequences ABU69126-
 CC ABU69171 represent human NOVX polypeptides of the invention
 XX
 SQ Sequence 319 AA;
 Query Match 98.9%; Score 1718.5; DB 6; Length 319;
 Best Local Similarity 98.7%; Pred. No. 1.3e-152;
 Matches 315; Conservative 0; Mismatches 1; Indels 3; Gaps 1;
 QY 1 MTRTYENFOYLENKVKVQGFKNGLPLQSLQRLRSGPCHLLSLGLGLLLVLCVVG 60
 DB 1 MTRTYENFOYLENKVKVQGFKNGLPLQSLQRLRSGPCHLLSLGLGLLLVLCVVG 60
 QY 61 QNSKFORDLVLTDFSNFTSNFTVAETIQTALTSQSSLEETIASLKAEVEGFKQERQAGVS 120
 DB 61 QNSKFORDLVLTDFSNFTSNFTVAETIQTALTSQSSLEETIASLKAEVEGFKQERQAGVS 120
 QY 121 ELQEHHTQKAHLGCHPCPSVCPVHSEMMLRVQDLVKLTCCVATLNNN--ASTE 177
 DB 121 ELQEHHTQKAHLGCHPCPSVCPVHSEMMLRVQDLVKLTCCVATLNNGEEASTE 180
 QY 178 GTCCPVNVWEHQDSYWFHSHGMSWAEAEKYCOLKNAHLVIVINSRERQNFVKYLGSAYT 237
 DB 181 GTCCPVNVWEHQDSYWFHSHGMSWAEAEKYCOLKNAHLVIVINSRERQNFVKYLGSAYT 240
 QY 238 WMGLSDPEGAKWVDGTDYATGPNWKPQDDMQHGLGGGDCAHFHPDGRWDDVCQ 297
 DB 241 WMGLSDPEGAKWVDGTDYATGPNWKPQDDMQHGLGGGDCAHFHPDGRWDDVCQ 300
 QY 298 RPYHWVCEAGLGQTSQESH 316
 DB 301 RPYHWVCEAGLGQTSQESH 319
 RESULT 4
 ADO08343
 ID ADO08343 standard; protein; 315 AA.
 XX
 AC ADO08343;
 XX
 DT 01-JUL-2004 (first entry)
 XX
 DE Human NOVX polypeptide #45.
 XX
 KW Human; NOVX; cardiomyopathy; atherosclerosis; hypertension; scleroderma;
 KW obesity; cancer; diabetes; haemophilia; graft-versus-host disease; AIDS;
 KW asthma; Crohn's disease; multiple sclerosis; infection; anorexia;
 KW cancer-associated cachexia; neurodegenerative disorder;
 KW Alzheimer's disease; Parkinson's disease; haematopoietic disorder;
 KW wasting disorder.
 XX
 OS Homo sapiens.
 XX
 PN US2004018594-A1.
 XX
 PD 29-JAN-2004.
 XX
 PD 01-MAY-2002; 2002US-00138588.
 PF
 XX 03-MAY-2001; 2001US-0288395P.
 PR 04-MAY-2001; 2001US-0288900P.
 PR 07-MAY-2001; 2001US-0289087P.
 PR 14-MAY-2001; 2001US-0290753P.
 PR 15-MAY-2001; 2001US-0291189P.
 PR 16-MAY-2001; 2001US-0291243P.
 PR 18-MAY-2001; 2001US-0292001P.
 PR 21-MAY-2001; 2001US-0292374P.
 PR 22-MAY-2001; 2001US-0292587P.
 PR 23-MAY-2001; 2001US-0293107P.
 PR 29-MAY-2001; 2001US-0294110P.
 PR 30-MAY-2001; 2001US-0294434P.

PR 31-MAY-2001; 2001US-0294827P.
 PR 31-JUL-2001; 2001US-0308901P.
 PR 17-AUG-2001; 2001US-0313388P.
 PR 21-AUG-2001; 2001US-0313851P.
 PR 21-AUG-2001; 2001US-0313937P.
 PR 17-SEP-2001; 2001US-0322701P.
 PR 17-SEP-2001; 2001US-0322802P.
 PR 25-SEP-2001; 2001US-0324757P.
 PR 27-SEP-2001; 2001US-0325314P.
 PR 27-SEP-2001; 2001US-0325682P.
 PR 01-NOV-2001; 2001US-0332129P.
 PR 03-DEC-2001; 2001US-0336882P.
 PR 14-DEC-2001; 2001US-0340305P.
 XX
 PA (ALSO/) ALSOBROOK J P.
 PA (ANDE/) ANDERSON D W.
 PA (BOLD/) BOLDOG F L.
 PA (BURG/) BURGESS C E.
 PA (CASW/) CASMAN S J.
 PA (CHAP/) CHAPOVAL A.
 PA (EDIN/) EDINGER S R.
 PA (GERL/) GERLACH V.
 PA (GORM/) GORMAN L.
 PA (GUNT/) GUNTHER E.
 PA (GUOX/) GUO X S.
 PA (KEKU/) KERUDA R.
 PA (LEPL/) LEPLEY D M.
 PA (LILL/) LI L.
 PA (LIUX/) LIU X.
 PA (MALY/) MALYANKAR U M.
 PA (MILL/) MILLER C E.
 PA (MILL/) MILLER I.
 PA (PADI/) PADIGARU M.
 PA (PATT/) PATTURAJAN M.
 PA (PENA/) PENA C E A.
 PA (RIEG/) RIEGER D K.
 PA (SHEN/) SHENOY S G.
 PA (SHIM/) SHIMKETS R A.
 PA (SPYT/) SPYTEK K A.
 PA (TAUP/) TAUPIER R J.
 PA (VERN/) VERNET C A M.
 PA (VOSS/) VOSS E Z.
 PA (ZERH/) ZERHUSEN B D.
 XX
 PI Alsbrook JP, Anderson DW, Boldog FL, Burgess CE, Casman SJ;
 PI Chapoval A, Edinger SR, Gerlach V, Gorman L, Gunther E, Guo XS;
 PI Kekuda R, Lepley DM, Li L, Liu X, Malyankar UM, Miller CE;
 PI Millet I, Padigaru M, Patturajan M, Pena CE, Rieger DK, Shenoy SG;
 PI Shimkets RA, Spytek KA, Taupier RJ, Vernet CAM, Voss EZ;
 PI Zerhusen BD;
 XX
 WIPI: 2004-122037/12.
 DR N-PSDB; ADO08342.
 XX
 PT New NOVX polypeptides and nucleic acids, useful for preventing or
 PT treating NOVX-associated disorders, e.g. cancer, diabetes,
 PT atherosclerosis, asthma or AIDS, and in chromosome mapping, tissue typing
 PT or pharmacogenomics.
 XX
 PS Claim 2; SEQ ID NO 90; 219pp; English.
 XX
 CC The invention relates to human NOVX polypeptides and the polynucleotides
 CC encoding them. The polypeptides, polynucleotides and antibodies that bind
 CC immunospecifically to the polypeptides are useful in diagnosing, treating
 CC or preventing NOVX-associated disorders such as cardiomyopathy,
 CC atherosclerosis, hypertension, scleroderma, obesity, cancer, diabetes,
 CC haemophilia, graft-versus-host disease, AIDS, asthma, Crohn's disease,
 CC multiple sclerosis, infections, anorexia, cancer-associated cachexia,
 CC neurodegenerative disorders (e.g. Alzheimer's disease or Parkinson's
 CC disease), haematopoietic disorders and wasting disorders. The
 CC polynucleotides are also used as hybridisation probes, in chromosome
 CC mapping and in tissue typing. The polypeptides are also useful as
 CC vaccines. This sequence represents a human NOVX polypeptide of the

CC invention.
 XX
 SQ Sequence 315 AA;
 Query Match 95.0%; Score 1651.5; DB 8; Length 315;
 Best Local Similarity 96.9%; Pred. No. 2.6e-146;
 Matches 308; Conservative 0; Mismatches 5; Indels 5; Gaps 3;
 QY 1 MTRTYENFOYLENKVKVQGFKNKGNPLPLQSLQRLRSRSGPCHLLLSLGLGLLLVLCVVG 60
 DB 1 MTRTYENFOYLENKVKVQGFKNKGNPLPLQSLQRLRSRSGPCHLLLSLGLGLLLVLCVVG 60
 QY 61 QNSKFORDLVLRDTSNFTSNFTVAEIQALTSQSSLEETIASLKAEEVGFQKQRQAGVS 120
 DB 61 QNSKFORDLVLRDTSNFTSNFTVAEIQALTSQSSLEETIASLK- -VEGFQKQRQAGVS 118
 QY 121 ELOHHTTQKAHLGHCPHCPSCVCPVHSEMLRVOQLVODLKLTCQVATIAN- -NASTEG 178
 DB 119 ELOHHTTQK- -HLGHCPHCPSCVCPVHSEMLRVOQLVODLKLTCQVATLNNGEASTEG 177
 QY 179 TCCPVNWEHODSCYWFSSHSGMSWAEAEKYCOLKNAHLVWINSREEQNFVKYLGSAATW 238
 DB 178 TCCPVNWEHODSCYWFSSHSGMSWAEAEKYCOLKNAHLVWINSREEQNFVKYLGSAATW 237
 QY 239 MGLSDPEGAWKWDGTDYATQFNWKPGQDDWQGHGLGGGEDCAHFHPDGRWDDVCOR 298
 DB 238 MGLSDPEGAWKWDGTDYATQFNWKPGQDDWQGHGLGGGEDCAHFHPDGRWDDVCOR 297
 QY 299 PYHWVCEAGLGTQSQESH 316
 DB 298 PYHQVCEAGKGTQSQESH 315
 RESULT 5
 AAW88129
 ID AAW88129 standard; protein; 273 AA.
 XX
 AC AAW88129;
 XX
 DT 11-MAY-1999 (first entry)
 DE Variant primate DCM2 C-lectin family gene protein sequence.
 XX
 KW Primate; dendritic cell membrane protein; DCMPI; DCM2;
 KW chromosomal abnormality; expression misregulation;
 KW abnormal proliferation; regeneration; degeneration; haematopoietic cell.
 XX
 OS Mammalia.
 XX
 PN WO9902562-A1.
 XX
 PD 21-JAN-1999.
 XX
 PF 08-JUL-1998; 98WO-US013436.
 XX
 PR 09-JUL-1997; 97US-0053080P.
 XX
 PA (SCHE) SCHERING CORP.
 XX
 PI Valladeau J, Ravel O, Bates EEM, Ford J, Saeland S, Lebecque SJE;
 XX
 WIPI: 1999-120786/10.
 DR N-PSDB; AAX04868.
 XX
 PT Dendritic cell membrane proteins - used to treat conditions associated
 PT with abnormal physiology or development.
 XX
 PS Claim 2(b); Page 76-78; 82pp; English.
 XX
 CC Dendritic cell membrane protein 1 (DCMP1) and DCM2 nucleic acids can be
 CC used as markers for distinguishing cell types, including genomic aspects
 CC of cells, as well as mRNA and protein expression patterns. They can also
 CC be used to detect chromosomal abnormalities. The proteins can be used to

CC diagnose disorders associated with expression misregulation. They can
 CC also be used to treat conditions associated with abnormal physiology or
 CC development, including abnormal proliferation, e.g. cancerous conditions
 CC or degenerative conditions. Abnormal proliferation, regeneration,
 CC degeneration and atrophy may be modulated using the proteins. The
 CC proteins may also play a role in regulation or development of
 CC haematopoietic cells
 XX
 SQ Sequence 273 AA;

Query Match 82.1%; Score 1427.5; DB 2; Length 273;
 Best Local Similarity 84.6%; Pred. No. 2.4e-125;
 Matches 270; Conservative 0; Mismatches 0; Indels 49; Gaps 3;

QY 1 MTRTYENFQYLENKVKVQGFNGPLPLQSLQLRLSGPCHLLLSGLGLLLVLCVVG 60
 DQ 1 MTRTYENFQYLENKVKVQGFNGPLPLQSLQLRLSGPCHLLLSGLGLLLVLCVVG 41
 QY 61 QNSKFORDLVTLRTDFSNFTSNTVAEIQALTSQSSLEETIASLKAEVEGFKQERQAGVS 120
 DQ 42 QNSKFORDLVTLRTDFSNFTSNTVAEIQALTSQSSLEETIASLKAEVEGFKQERQAGVS 98
 QY 121 ELQEHHTTQKAHLGHCPCPCVCPVHSEMLLRVQQLVQDLKCLTCQVATLNNN---ASTE 177
 DQ 99 -----VHSEMLLRVQQLVQDLKCLTCQVATLNNNGEERASTE 134
 QY 178 GTCCPVNWEHQDSYWFHSGMSWABAEKYCOLKNAHLVVIINSRBNFQVQKYLGSAYT 237
 DQ 135 GTCCPVNWEHQDSYWFHSGMSWABAEKYCOLKNAHLVVIINSRBNFQVQKYLGSAYT 194
 QY 238 WMGLSDPEGAWKVVDGTDYATGFQNKWPGQDDWQGHGLGGEDCAHFHDPGRWDDVCQ 297
 DQ 195 WMGLSDPEGAWKVVDGTDYATGFQNKWPGQDDWQGHGLGGEDCAHFHDPGRWDDVCQ 254
 QY 298 RPYHWVCEAGLGQTSQESH 316
 DQ 255 RPYHWVCEAGLGQTSQESH 273

RESULT 6
 ABU69169
 ID ABU69169 standard; protein; 267 AA.
 XX
 AC ABU69169;
 XX
 DT 02-JUN-2003 (first entry)
 XX
 DE Human NOVX polypeptide #44.
 XX
 KW Human; NOVX; metabolic disorder; diabetes; infectious disease; obesity;
 KW anorexia; cancer; cardiovascular disorder; asthma; neurogenesis;
 KW neurodegenerative disorder; epilepsy; immune disorder; osteoarthritis;
 KW haematopoietic disorder; inflammatory skin disorder; dyslipidemia;
 KW haematopoiesis; wound healing; angiogenesis; bacterial infection;
 KW viral infection; fungal infection; helminthic infection; atherosclerosis;
 KW protozoal infection; hypertension.
 XX
 OS Homo sapiens.
 XX
 PN W0200290504-A2.
 XX
 PD 14-NOV-2002.
 XX
 XX 02-MAY-2002; 2002WO-US014342.
 XX
 XX 03-MAY-2001; 2001US-0288395P.
 XX
 XX 07-MAY-2001; 2001US-0288900P.
 XX
 XX 14-MAY-2001; 2001US-0290753P.
 XX
 XX 15-MAY-2001; 2001US-0291189P.
 XX
 XX 16-MAY-2001; 2001US-0291243P.
 XX
 XX 18-MAY-2001; 2001US-0292041P.
 XX
 XX 21-MAY-2001; 2001US-0292374P.

PR 22-MAY-2001; 2001US-0292587P.
 PR 23-MAY-2001; 2001US-0293107P.
 PR 29-MAY-2001; 2001US-0294110P.
 PR 30-MAY-2001; 2001US-0294434P.
 PR 31-MAY-2001; 2001US-0294827P.
 PR 18-JUN-2001; 2001US-0298988P.
 PR 31-JUL-2001; 2001US-0308901P.
 PR 17-AUG-2001; 2001US-0313388P.
 PR 21-AUG-2001; 2001US-0313851P.
 PR 21-AUG-2001; 2001US-0313937P.
 PR 17-SEP-2001; 2001US-0322701P.
 PR 17-SEP-2001; 2001US-0322802P.
 PR 25-SEP-2001; 2001US-0324757P.
 PR 27-SEP-2001; 2001US-0325314P.
 PR 27-SEP-2001; 2001US-0325682P.
 PR 21-NOV-2001; 2001US-0332129P.
 PR 03-DEC-2001; 2001US-0336882P.
 PR 14-DEC-2001; 2001US-0340305P.
 PR 01-MAY-2002; 2002US-00138588.
 XX
 XX (CURA-) CURAGEN CORP.
 XX
 PI Alsbrook JP, Anderson DW, Boldog FL, Burgess CE, Casman SJ;
 PI Chapoval A, Edinger S, Gerlach V, Gorman L, Gunther E, Guo X;
 PI Kekuda R, Lepley DM, Li L, Liu X, Malyankar UM, Miller CE;
 PI Milliet I, Padigaru M, Patturajan M, Pena CEA, Rieger DK, Shenoy SG;
 PI Shimkets RA, Spytek KA, Taupier RJ, Vernet CAM, Voss EZ;
 PI Zerhusen BD;
 XX
 XX WPI; 2003-103512/09.
 DR N-PSDB; ACA10154.
 XX
 PT New isolated NOVX polypeptides and polynucleotides, useful for
 PT preventing, diagnosing or treating NOVX-associated disorders, e.g.
 PT osteoarthritis, obesity, atherosclerosis, cancer, Parkinson's disease,
 PT asthma, or infections.
 XX
 PS Claim 2; Page 169; 340pp; English.
 XX
 CC The invention relates to human NOVX polypeptides and the polynucleotides
 CC encoding them. The polypeptides, polynucleotides and antibodies that bind
 CC immunospecifically to the polypeptides are useful in the manufacture of a
 CC medicament for treating a syndrome associated with a human disease,
 CC preferably a NOVX-associated disorder. The sequences are useful for
 CC treating, preventing or diagnosing diseases such as metabolic disorders,
 CC diabetes, obesity, infectious diseases (viral, bacterial, fungal,
 CC helminthic, and protozoal), anorexia, cancer, cardiovascular disorders
 CC (e.g. hypertension, atherosclerosis), neurodegenerative disorders (e.g.
 CC Alzheimer's disease, Parkinson's disease), epilepsy, immune disorders,
 CC osteoarthritis, haematopoietic disorders, inflammatory skin disorders,
 CC asthma and various dyslipidemias. The nucleic acids and polypeptides may
 CC also be used as targets for the identification of small molecules that
 CC modulate or inhibit e.g. neurogenesis, cell differentiation, cell
 CC proliferation, haematopoiesis, wound healing and angiogenesis, and in the
 CC generation of antibodies that bind immunospecifically to NOVX substances
 CC for use in therapeutic or diagnostic methods. The nucleic acids are
 CC further used as hybridisation probes, and in chromosome mapping, tissue
 CC typing, preventive medicine and pharmacogenomics. Sequences ABU69126-
 CC ABU69171 represent human NOVX polypeptides of the invention
 XX
 SQ Sequence 267 AA;

Query Match 81.2%; Score 1410.5; DB 6; Length 267;
 Best Local Similarity 89.5%; Pred. No. 9.1e-124;
 Matches 263; Conservative 0; Mismatches 4; Indels 27; Gaps 1;

QY 23 GPLPLQSLQLRLSGPCHLLLSGLGLLLVLCVVGQNSKFORDLVTLRTDFSNFTSN 82
 DQ 1 GPLPLQSLQLRLSGPCHLLLSGLGLLLVLCVVGQNSKFORDLVTLRTDFSNFTSN 60
 QY 83 TVAEIQALTSQSSLEETIASLKAEVEGFKQERQAGVSLEQHTTQKAHLGHCPCPCV 142
 DQ 61 TVAEIQALTSQSSLEETIASLKAEVEGFKQERQAGVSLEQHTTQKAHLGHCPCPCV 95

QY 143 VPHSEMLLRVQQLVQDLKLTCCQVATLNNASTEGTCCPVNVVBEHQDSCYWFSGMSW 202
DB 96 --VHSEMLLRVQQLVQDLKLTCCQVATLNNASTEGTCCPVNVVBEHQDSCYWFSGMSW 153
QY 203 ABAEKYCOLKNAHLVVINSREONFVQKYLGSAYTWMGLSDPEGAKWVDGTDYATGFON 262
DB 154 ABAEKYCOLKNAHLVVINSREONFVQKYLGSAYTWMGLSDPEGAKWVDGTDYATGFON 213
QY 263 WKPQDDQHGGLGGEDCAHFHPDGRWDDVCQRPYHVCENGIGOTSQESH 316
DB 214 WKPQDDQHGGLGGEDCAHFHPVGRWDDVCQRPYHVCENGIGOTSQESH 267

RESULT 7
ADO08341
ID ADO08341 standard; protein; 267 AA.
AC ADO08341;
XX
XX 01-JUL-2004 (first entry)
XX Human NOVX polypeptide #44.
DE
XX Human; NOVX; cardiomyopathy; atherosclerosis; hypertension; scleroderma;
KW obesity; cancer; diabetes; haemophilia; graft-versus-host disease; AIDS;
KW asthma; Crohn's disease; multiple sclerosis; infection; anorexia;
KW cancer-associated cachexia; neurodegenerative disorder;
KW Alzheimer's disease; Parkinson's disease; haematopoietic disorder;
KW wasting disorder.
XX
XX Homo sapiens.
XX
XX US2004018594-A1.
XX
XX 29-JAN-2004.
XX
XX 01-MAY-2002; 2002US-00138588.
XX
XX 03-MAY-2001; 2001US-0288395P.
XX
XX 04-MAY-2001; 2001US-0288900P.
XX
XX 07-MAY-2001; 2001US-0289087P.
XX
XX 14-MAY-2001; 2001US-0290753P.
XX
XX 15-MAY-2001; 2001US-0291189P.
XX
XX 16-MAY-2001; 2001US-0291243P.
XX
XX 18-MAY-2001; 2001US-0292001P.
XX
XX 21-MAY-2001; 2001US-0292374P.
XX
XX 22-MAY-2001; 2001US-0292587P.
XX
XX 23-MAY-2001; 2001US-0293107P.
XX
XX 29-MAY-2001; 2001US-0294110P.
XX
XX 30-MAY-2001; 2001US-0294434P.
XX
XX 31-MAY-2001; 2001US-0294827P.
XX
XX 31-JUL-2001; 2001US-0308901P.
XX
XX 17-AUG-2001; 2001US-0313388P.
XX
XX 21-AUG-2001; 2001US-0313851P.
XX
XX 21-AUG-2001; 2001US-0313937P.
XX
XX 17-SEP-2001; 2001US-0322701P.
XX
XX 17-SEP-2001; 2001US-0322802P.
XX
XX 25-SEP-2001; 2001US-0324757P.
XX
XX 27-SEP-2001; 2001US-0325314P.
XX
XX 27-SEP-2001; 2001US-0325682P.
XX
XX 21-NOV-2001; 2001US-0332129P.
XX
XX 03-DEC-2001; 2001US-0336882P.
XX
XX 14-DEC-2001; 2001US-0340305P.
XX
XX (ALSO/) ALSOBROOK J P.
PA (ANDE/) ANDERSON D W.
PA (BOLD/) BOLDOGF F L.
PA (BURG/) BURGESS C E.
PA (CASM/) CASMAN S J.
PA (CHAP/) CHAFOVAL A.
PA (EDIN/) EDINGER S R.
PA (GERL/) GERLACH V.

PA (GORM/) GORMAN L.
PA (GUNT/) GUNTHER E.
PA (GUOX/) GUO X S.
PA (KEKU/) KEKUDA R.
PA (LEPL/) LEFLEY D M.
PA (LILL/) LI L.
PA (LITX/) LIU X.
PA (MALY/) MALYANKAR U M.
PA (MILL/) MILLER C E.
PA (MILL/) MILLET I.
PA (PADI/) PADIGARU M.
PA (PATT/) PATTURAJAN M.
PA (PENA/) PENA C E A.
PA (RIEG/) RIEGER D K.
PA (SHEN/) SHENOY S G.
PA (SHIM/) SHIMKETS R A.
PA (SPYT/) SPYTEK K A.
PA (TAUP/) TAUPIER R J.
PA (VERN/) VERNET C A M.
PA (VOSS/) VOSS E Z.
PA (ZERH/) ZERHUSEN B D.
XX
XX Alsbrook JP, Anderson DW, Boldog FL, Burgees CE, Caaman SJ;
PI Chapoval A, Edinger SR, Gerlach V, Gorman L, Gunther E, Guo XS;
PI Kekuda R, Lepley DM, Li L, Liu X, Malyankar UM, Miller CE;
PI Millet I, Padigar M, Patturajan M, Pena CEA, Rieger DK, Shenoy SG;
PI Shmkets RA, Spytek KA, Taupier RJ, Vernet CAM, Voss EZ;
PI Zerhusen BD;
XX
XX WPI; 2004-122037/12.
DR N-PSDB; ADO08340.
XX
XX New NOVX polypeptides and nucleic acids, useful for preventing or
PT treating NOVX-associated disorders, e.g. cancer, diabetes,
PT atherosclerosis, asthma or AIDS, and in chromosome mapping, tissue typing
PT or pharmacogenomics.
XX
XX Claim 2; SEQ ID NO 88; 219pp; English.
XX
XX The invention relates to human NOVX polypeptides and the polynucleotides
CC encoding them. The polypeptides, polynucleotides and antibodies that bind
CC immunospecifically to the polypeptides are useful in diagnosing, treating
CC or preventing NOVX-associated disorders such as cardiomyopathy,
CC atherosclerosis, hypertension, scleroderma, obesity, cancer, diabetes,
CC haemophilia, graft-versus-host disease, AIDS, asthma, Crohn's disease,
CC multiple sclerosis, infections, anorexia, cancer-associated cachexia,
CC neurodegenerative disorders (e.g. Alzheimer's disease or Parkinson's
CC disease), haematopoietic disorders and wasting disorders. The
CC polynucleotides are also used as hybridisation probes, in chromosome
CC mapping and in tissue typing. The polypeptides are also useful as
CC vaccines. This sequence represents a human NOVX polypeptide of the
CC invention.
XX
XX Sequence 267 AA;
SQ
Query Match 81.2%; Score 1410.5; DB 8; Length 267;
Best Local Similarity 89.5%; Pred. No. 9.1e-124;
Matches 263; Conservative 0; Mismatches 4; Indels 27; Gaps 1;
QY 23 GPLPLQSLQLRLRGPCCHLLSLGLGILLVLCVGFQNSKFORDLVTLRTDTSNFTSN 82
DB 1 GPLPLQSLQLRLRGPCCHLLSLGLGILLVLCVGFQNSKFORDLVTLRTDTSNFTSN 60
QY 83 TVAEIQALTSQGSLEETIASIKAEVGFQKORAGVSELOEHTTQKAHLGHCPCPSVC 142
DB 61 TVAEIQALTSQGSLEETIASIKAEVGFQKORQA----- 95
QY 143 VPHSEMLLRVQQLVQDLKLTCCQVATLNNASTEGTCCPVNVVBEHQDSCYWFSGMSW 202
DB 96 --VHSEMLLRVQQLVQDLKLTCCQVATLNNASTEGTCCPVNVVBEHQDSCYWFSGMSW 153
QY 203 ABAEKYCOLKNAHLVVINSREONFVQKYLGSAYTWMGLSDPEGAKWVDGTDYATGFON 262

Best Local Similarity 54.8%; Pred. No. 6.3e-78;	
Matches 172; Conservative 47; Mismatches 67; Indels 29; Gaps 5;	
Qy	1 MTRTYENFOYLENKVK-VQGFKNQGPLPLOSLLQRLRSGPCHLLSLGLGLLLVLIICVWG 59
Db	1 MTKKEYDQHLNDNEEDHQLRKGPFPPLQRLCSGPRLLLSLGLSLLLVVVCVIG 60
Qy	60 FQNSKFRQDLVLTLRTDFSNFTSNVABIQALTSQGSLEETIASLKAEVGFQKQERQAGV 119
Db	61 SQNSQLQEELRLGURETFSNFTASTEAQVKGSLTQCGNVGRMKSLSQLSLE-KQOK---- 114
Qy	120 SELQEHHTQKAHLGHCPCPSVCVPVHSEMLLAVQQLVQDLKKLTQVATLNNASTECT 179
Db	115 -DLSED-----HSSLLLVHKQFVSDLRSLSCQMAALQNGS-ERT 152
Qy	180 CCPVNVVEHODSCYWFSGMSWAEAEKYCOLKNAHLVWINSREONFVKYLGSAVTWM 239
Db	153 CCPVNVVEHERSCYWFSGSKAWADADNYCRLEDAHLVWVTSWEEQKFVQHHGPPVNTWM 212
Qy	240 GLSDPEGAKWVDGTDYATGFQNWKFQPPDDWGHLGGGEDCAHFHPDGRWDDVCORP 299
Db	213 GLHDQNGPWKWDGTDYQTGFKNWRPEQPPDWYGHGLGGGEDCAHFHTDDGRWDDVCORP 272
Qy	300 YHWVCEAGLGQTSQE 314
Db	273 YRWVCEETLDSKASQE 287
RESULT 9	
ABM82572	ABM82572 standard; protein; 319 AA.
ID	ABM82572;
AC	ABM82572;
AC	
XX	
DT	18-NOV-2004 (first entry)
DB	Human diagnostic and therapeutic pprotein SEQ ID NO:2821.
XX	
KW	gene therapy; human diagnostic and therapeutic polynucleotide; dthcp.
XX	
OS	Homo sapiens.
XX	
PN	WO2004023973-A2.
XX	
PD	25-MAR-2004.
XX	
PF	12-SEP-2003; 2003WO-US028227.
XX	
PR	12-SEP-2002; 2002US-0410259P.
PR	12-SEP-2002; 2002US-0410260P.
XX	
PA	(INCY-) INCYTE CORP.
XX	
PI	Schmidt JP, Wright RU, Bruns CM, Marjanovic MM, Shen F;
PI	Harthshorne TA, Suchorolski MT, Altus CM, Pitts SJ, Elder LV;
PI	Mooney EM, Delegeane AM, Panesar IS, Banville SC, Reddy TP;
PI	Stevens KA, Blanchard JL, Panzer SR, Wang X, Au AP, Garstin EH;
PI	Peralta CH, Anderson SB, Rioux F, Shen EJ, Wu MC, Stuve LL;
PI	Lagace RE, Spiro PA, Stewart EA, Wingrove J, Vitt UA, Kirtan ES;
PI	Xu Y, Kwong M, Policky JL, Hurwitz BL, Ma Y, Jackson JL, Gietzen D;
PI	Patury S, Shi X, Suarez CJ;
XX	
DR	WPI; 2004-329368/30.
DR	N-PSDB; ACN41224.
XX	
PT	New diagnostic and therapeutic polynucleotides and polypeptides, useful
PT	in diagnosing a condition, disease or disorder associated with human
PT	molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or
PT	in gene mapping.
XX	
PS	Claim 27; Page; 190pp; English.

The invention relates to novel diagnostic and therapeutic polynucleotides


```
PA (INCY-) INCYTE CORP.
XX
PI Schmidt JP, Wright RJ, Bruns CM, Marjanovic MM, Shen F;
PI Harthorne TA, Suchorolski MT, Altus CM, Pitts SJ, Elder LV;
PI Mooney EM, Deleage AM, Panesar IS, Banville SC, Reddy TP;
PI Stevens KA, Blanchard JL, Panzer SR, Wang X, Au AP, Gerstin EH;
PI Paralta CH, Anderson SB, Rioux P, Shen EJ, Wu MC, Stuve LL;
PI Lagace RE, Spiro PA, Stewart EA, Wingrove J, Vitt UA, Kitton ES;
PI Xu Y, Kwong M, Policky JL, Hurwitz BL, Ma Y, Jackson JL, Gietzen D;
PI Patry S, Shi X, Suarez CJ;
XX
DR WPI; 2004-329368/30.
DR N-PSDB; ACN41222.
XX
XX New diagnostic and therapeutic polynucleotides and polypeptides, useful
PT in diagnosing a condition, disease or disorder associated with human
PT molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or
PT in gene mapping.
XX
XX Claim 27; Page; 190pp; English.
XX
XX The invention relates to novel diagnostic and therapeutic polynucleotides
CC selected from one of the 2722 sequences defined in the specification. A
CC polynucleotide of the invention may have a use in gene therapy. The human
CC diagnostic and therapeutic polynucleotides (dithp) or polypeptides may be
CC used to diagnose a particular condition, disease or disorder associated
CC with human molecules, e.g. cell proliferative disorders,
CC autoimmune/inflammatory disorder, developmental disorder, endocrine
CC disorder, neurological disorders, gastrointestinal disorders, or
CC infections caused by virus, bacteria, fungi or parasite. The dithp
CC molecules may also be used in genetic mapping, in identifying individuals
CC from minute biological samples, in detecting single nucleotide
CC polymorphisms, as molecular weight markers, and for somatic or germline
CC gene therapy. The present sequence represents a dithp protein of the
CC invention. Note: The sequence data for this patent is not represented in
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at www.wipo.int/pct/en/sequences/listing.htm
XX
XX Sequence 290 AA;
XX
Query Match 49.6%; Score 862; DB 8; Length 290;
Best Local Similarity 48.5%; Pred. No. 3.8e-72;
Matches 164; Conservative 37; Mismatches 61; Indels 76; Gaps 4;
QY 1 MTRTYENFOYLENKVK-VQGFKNGLPLQSLQLRLSGPCHLLSLGLGLLLVLCVWG 59
DB 1 MTRKEYQDLQHLNDESDHQLRKGPPLPQLQLCSGPRLLLSLGLLLVLCVWG 60
QY 60 FQNSKFORDLVLTIRDTFSNFTSNTVAIEIQALTSQGSLEETIASLKAEVGFQKQAGV 119
DB 61 SQNSQLEELRGLRETFNSFTASTEAQVKGSLTOD----- 95
QY 120 SELQHTTQKAHLGHCHPCPSVCVPVHSEMLLRVQQLVQDLKLTTCQVATLNNASTEGT 179
DB 96 -----HSLLLHVQFVSURLSUSCOMAALQNGS-ERT 128
QY 180 CCPVNWYEHODSCYWFHSGMSWAEKCYCOLKNAHLVWINSRBE----- 224
DB 129 CCPVNWYEHRSYCFWFSRSGKAMADNYCRLEDAHLVWVTSWEEQVTRTRVWEAGWPRR 188
QY 225 -----QNFVQKLGATYTWGLSDPEGAWKVVDGTATGPNQWKPQDPDQGHGL 276
DB 189 DHHPPLSLQKRFVQHHIGPVNTWMLHDQNGQPKWVDGTYETGPKNWRPQPDWYGHGL 248
QY 277 GGGEDCAHFHPDGRWDDVQCRPVHWCCEAGLQTSOE 314
DB 249 GGGEDCAHFTDDGRWDDVQCRPVHWCCEAGLQTSOE 286
RESULT 15
AAW15250
ID AAW15250 standard; protein; 274 AA.
XX
```

```
AC AAW15250;
XX
DT 17-OCT-2003 (revised)
DT 23-NOV-1997 (first entry)
XX
DE Asialoglycoprotein receptor H1 cytoplasmic+extracellular domains.
XX
KW Asialoglycoprotein receptor H1; AGPR; autoimmune hepatitis; autoantibody;
KW diagnosis.
XX
OS Homo sapiens.
OS Chimeric.
XX
XX Location/Qualifiers
FT Domain 1..42
FT /label= Cytoplasmic_domain
FT Domain 43..274
FT /label= Extracellular_domain
XX
XX EP773289-A2.
XX
XX 14-MAY-1997.
XX
XX 20-AUG-1996; 96EP-00113349.
XX
XX 21-AUG-1995; 95JP-00212118.
XX
XX (TOFU) TONEN CORP.
XX
XX Tanida E, Ohue C, Yagi S, Hasegawa A, Kiyosawa K, Yano A;
XX WPI; 1997-261316/24.
XX
XX Asialo:glyco:protein receptor H1 and L-H2 soluble derivatives - comprise
XX extracellular domains, optionally also with cytoplasmic domains, useful
XX for autoimmune hepatitis diagnosis.
XX
XX Claim 3; Page; 40pp; English.
XX
XX This polypeptide sequence comprises the cytoplasmic (CTD) and
XX extracellular (ECD) domains of human full-length asialoglycoprotein
XX receptor (AGPR) H1 (see AAW15245). It can be produced on a large scale,
XX in a form free of contaminating liver antigens, by PCR amplification (see
XX AAR6962-65) of CTD+ECD DNA (see also AAR6950), and expression in
XX bacterial (pref. E. coli) or animal (pref. mammalian) host cells. The ECD
XX alone (AAW15249) and AGPR L-H2 derivatives (see AAW15251-52) are also
XX claimed. The appearance of autoantibodies against AGPR can be used as an
XX indicator for autoimmune hepatitis (AIH), a disease that can lead to
XX cirrhosis and fatal intractable hepatitis. Recombinant soluble AGPR
XX derivatives can be used in a claimed method for detecting and measuring
XX anti-AGPR antibodies, e.g. by ELISA, Western blotting, etc. This allows
XX AIH to be diagnosed and distinguished from other diseases. AGPR
XX derivatives or full-length receptors can also be used for the analysis
XX and purification of asialoglycoproteins and to develop inhibiting agents
XX against asialoglycoprotein incorporation, or viral and other protein
XX invasion, into liver cells. (Updated on 17-OCT-2003 to standardise OS
XX field)
XX
XX Sequence 274 AA;
XX
Query Match 48.2%; Score 838; DB 2; Length 274;
Best Local Similarity 50.2%; Pred. No. 6.3e-70;
Matches 158; Conservative 46; Mismatches 65; Indels 46; Gaps 6;
QY 1 MTRTYENFOYLENKVK-VQGFKNGLPLQSLQLRLSGPCHLLSLGLGLLLVLCVWG 59
DB 1 MTRKEYQDLQHLNDESDHQLRKGPPLPQLQLCSGPRLLLSLGLGLLLVLCVWG 43
QY 60 FQNSKFORDLVLTIRDTFSNFTSNTVAIEIQALTSQGSLEETIASLKAEVGFQKQAGV 119
DB 44 SQNSQLEELRGLRETFNSFTASTEAQVKGSLTODGNGVGRKMSLESQLE--KQKQ--- 97
QY 120 SELQHTTQKAHLGHCHPCPSVCVPVHSEMLLRVQQLVQDLKLTTCQVATLNNASTEGT 179
```

